

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2003, 13:39:28 ; Search time 22.9677 Seconds  
(without alignments)  
371.305 Million cell updates/sec

Title: US-09-818-066-34\_COPY\_98\_161

Perfect score: 348

Sequence: 1 EAFRYQERPPETTPPS.....PLYQSEPAVPVITPPPLKKK 64

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_101002.\*  
1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	348	100.0	161	AAW11854	Duck hepatitis B v
2	83.5	24.0	906	AAW25681	Transgenic mouse N
3	83.5	24.0	907	AAW49032	Human eIF4G-like p
4	83.5	24.0	907	AB957158	Mouse ischaemic co
5	83.5	24.0	940	AAW71370	Death associated p
6	83.5	24.0	940	AAW58970	Breast and ovarian
7	77	22.1	1190	22	Drosophila melanog
8	76	21.8	280	14	Insecticidal/Fungi
9	76	21.8	708	23	Herbicidally activ
10	74.5	21.4	223	22	Human peptide #121

11	74.5	21.4	223	22	AB932621	Peptide #127 encod
12	74.5	21.4	223	22	AB918119	Protein #118 encod
13	74.5	21.4	223	22	AAW53451	Human brain expres
14	74.5	21.4	223	22	AAW65828	Human bone marrow
15	74.5	21.4	223	22	AAW13690	Peptide #124 encod
16	74.5	21.4	223	22	AAW26090	Peptide #127 encod
17	74.5	21.4	223	22	AAW01440	Peptide #122 encod
18	74.5	21.4	223	23	ABG35462	Human peptide enco
19	74.5	21.4	1004	19	AAW37856	Human polyhomeotic
20	74.5	21.4	1004	19	AAW52830	A tumour suppresso
21	74	21.3	531	22	ABG21540	Novel human diago
22	73.5	21.1	1042	22	ABG61265	Drosophila melanog
23	73	21.0	109	21	AAW64587	Nonclassical cadhe
24	73	21.0	544	21	AAW15453	Arabidopsis thalia
25	73	21.0	652	23	ABW92424	Herbicidally activ
26	73	21.0	794	18	AAW25637	Human cadherin-12
27	73	21.0	794	18	AAW13135	Putative human cad
28	73	21.0	849	22	ABG05943	Novel human diago
29	72.5	20.8	783	21	AAW23648	Murine pAPE protei
30	72.5	20.8	886	9	AAW0345	Sequence encoded b
31	72.5	20.8	886	23	ABP35643	Fungal ZBC protein
32	72.5	20.8	2665	22	ABW28314	Human peptide #965
33	72.5	20.8	2665	22	ABW33490	Peptide #996 encod
34	72.5	20.8	2665	22	ABW18950	Protein #949 encod
35	72.5	20.8	2665	22	AAW54270	Human brain expres
36	72.5	20.8	2665	22	AAW66665	Human bone marrow
37	72.5	20.8	2665	22	AAW14533	Peptide #967 encod
38	72.5	20.8	2665	22	AAW26950	Peptide #987 encod
39	72.5	20.8	2665	22	AAW02259	Peptide #941 encod
40	72.5	20.8	2665	23	ABW36319	Human peptide enco
41	72.5	20.8	3266	21	AAW42491	Human ORFX ORF2255
42	71.5	20.5	1047	22	ABW70874	Drosophila melanog
43	71	20.4	785	23	ABW97543	Novel human protei
44	70.5	20.3	443	22	ABW78130	Human actin 49
45	70.5	20.3	920	18	AAW25716	Mouse beta meltrin

ALIGNMENTS

RESULT 1  
AAW11854  
ID AAW11854 standard; Protein; 161 AA.  
AC AAW11854;  
XX 10-MAY-1997 (first entry)  
DT Duck hepatitis B virus pre-S domain.  
DE Hepadnavirus receptor; p120; p170; vaccine; pre-S domain.  
XX Hepadnavirus receptor; p120; p170; vaccine; pre-S domain.  
XX Duck hepatitis B virus.  
XX

PH	Key	Location/Qualifiers
FT	Binding-site	87..102
FT	/label=	p170_binding_site
FT	/note=	"the p170 binding site has been mapped to a major neutralising epitope of the pre-S domain (aa87-102), within which the Lys-95 and Arg-97 residues required for virion-receptor interaction"
FT	Peptide	1..102
FT	/note=	"claimed pre-S fragment (Claim 20) capable of binding hepadnavirus receptor"
FT	Peptide	1..104
FT	/note=	"claimed pre-S fragment (Claim 20) capable of binding hepadnavirus receptor"
FT	Peptide	1..126
FT	/note=	"claimed pre-S fragment (Claim 20) capable of binding hepadnavirus receptor"
FT	Peptide	1..138
FT	/note=	"claimed pre-S fragment (Claim 20) capable of binding hepadnavirus receptor"

FT Peptide 25..102 capable of binding hepadnavirus receptor"  
FT /note= "claimed pre-S fragment (Claim 20)  
FT capable of binding hepadnavirus receptor"  
FT Peptide 25..104  
FT /note= "claimed pre-S fragment (Claim 20)  
FT capable of binding hepadnavirus receptor"  
FT Peptide 25..126  
FT /note= "claimed pre-S fragment (Claim 20)  
FT capable of binding hepadnavirus receptor"  
FT Peptide 25..161  
FT /note= "claimed pre-S fragment (Claim 20)  
FT capable of binding hepadnavirus receptor"  
FT Peptide 42..102  
FT /note= "claimed pre-S fragment (Claim 20)  
FT capable of binding hepadnavirus receptor"  
FT Peptide 59..104  
FT /note= "claimed pre-S fragment (Claim 20)  
FT capable of binding hepadnavirus receptor"  
FT Peptide 59..126  
FT /note= "claimed pre-S fragment (Claim 20)  
FT capable of binding hepadnavirus receptor"  
FT Peptide 59..161  
FT /note= "claimed pre-S fragment (Claim 20)  
FT capable of binding hepadnavirus receptor"  
FT Peptide 71..161  
FT /note= "claimed pre-S fragment (Claim 20)  
FT capable of binding hepadnavirus receptor"  
FT Peptide 71..126  
FT /note= "claimed pre-S fragment (Claim 20)  
FT capable of binding hepadnavirus receptor"  
FT Peptide 80..104  
FT /note= "claimed pre-S fragment (Claim 20)  
FT capable of binding hepadnavirus receptor"  
FT Peptide 80..161  
FT /note= "claimed pre-S fragment (Claim 20)  
FT capable of binding hepadnavirus receptor"  
FT Peptide 87..161  
FT /note= "claimed pre-S fragment (Claim 20)  
FT capable of binding hepadnavirus receptor"  
FT Peptide 87..104  
FT /note= "claimed pre-S fragment (Claim 21)  
FT capable of binding hepadnavirus receptor"  
XX WO9704000-A1.  
XX  
XX  
XX 06-FEB-1997.  
XX  
XX 22-JUL-1996; 96WO-US12098.  
XX  
XX 21-JUL-1995; 95US-0001371.  
XX (GEO ) GEN HOSPITAL CORP.  
XX  
XX Li J, Tong S, Wands JR;  
XX WPI; 1997-132572/12.  
XX DR N-PSDB; AAT59583.  
XX  
XX p170, a new avian hepadnavirus receptor - binds to pre-S domain of  
XX duck hepatitis B virus, attenuated p170 may be used to immunise  
XX animals against hepadnaviral infection  
XX  
XX Claim 19; Page 87-88; 175pp; English.  
XX  
XX The pre-S domain (AAW11854) of the duck hepatitis B virus (DHBV)  
XX large envelope protein binds to the cellular p170 receptor (see  
XX also AAW11851-52) at a major neutralising epitope, within which are 2  
XX basic amino acids (Lys-95, Arg-97) required for virion-receptor  
XX interaction. Peptide sequences encompassing Lys-95 and Arg-97  
XX can act as p170 ligands. Such peptides can be used to reduce  
XX the level of hepadnaviral infection in an animal. The pre-S  
XX domain (esp. attenuated), and nucleic acids encoding it, can also

CC be used as vaccines to immunise animals against hepadnavirus  
CC infection.  
XX  
XX SQ Sequence 161 AA;  
Query Match 100.0%; Score 348; DB 18; Length 161;  
Best Local Similarity 100.0%; Pred. No. 5.4e-30;  
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 EAFRRYQERPEPTTTPSSPPQKLPQDDPLLGNSLLETHLYQSEPAVPVKTTP 60  
|||||  
Db 98 EAFRRYQERPEPTTTPSSPPQKLPQDDPLLGNSLLETHLYQSEPAVPVKTTP 157  
Qy 61 LKKK 64  
|||||  
Db 158 LKKK 161  
RESULT 2  
AAW25681  
ID AAW25681 standard; Protein; 906 AA.  
XX  
XX AC AAW25681;  
XX  
XX DT 31-MAR-1998 (first entry)  
XX  
XX Transgenic mouse NTAl polypeptide.  
DE  
XX RNA editing; mooring primer; screening; detection; editing enzyme;  
KW obesity; APOBEC-1; therapeutic; NTAl; transgenic.  
XX  
XX OS Synthetic.  
OS Homo sapiens.  
OS Mus sp.  
FH Key Location/Qualifiers  
FT Protein 1..906  
FT /label= NTAl  
FT /note= "partial coding sequence"  
XX WO9732032-A1.  
XX  
XX PD 04-SEP-1997.  
XX  
XX PF 28-FEB-1997; 97WO-US03189.  
XX  
XX PR 01-MAR-1996; 96US-0609230.  
XX  
XX (RECC ) UNIV CALIFORNIA.  
XX  
XX PI Innerarity TL, Qian X, Yamanaka S;  
XX WPI; 1997-448694/41.  
XX DR N-PSDB; AAT86087.  
XX  
XX Detection of mRNA targets for editing enzymes - by amplifying RNA  
XX from tissues by RT-PCR using mooring primers to identify genes  
XX responsible for non-wild type phenotype(s)  
XX  
XX Example 4; Figure 5; 36pp; English.  
XX  
XX This sequence represents a novel target of the apoB mRNA-editing  
XX enzyme catalytic polypeptide #1, APOBEC-1, (i.e NTAl) which is used  
XX as an example of a novel method of detecting mRNA candidates for editing  
XX in a tissue. The method involves the generation of cDNA from RNA in the  
XX tissue using one or more mooring primers and reverse transcriptase (RT)  
XX and is amplified by PCR using one or more mooring primers and a 5'  
XX arbitrary primer. The method has been used for identification of an  
XX obesity gene e.g. APOBEC-1 or an oncogene in an animal and also for the  
XX identification of non-wild type phenotypes. The identification of such  
XX mRNA's provides insight into the role of the genes encoding them, in  
XX growth, differentiation, or lipid accumulation, and also into the  
XX possible role of an RNA editing enzyme or polypeptide in causing obesity

Query Match 24.0%; Score 83.5; DB 19; Length 907;

AW71370

KW		antibacterial; antifungal; antiparasitic; cardiast; immune disorder;
KW	Addison's disease; allergy; autoimmune haemolytic anaemia;	
KW	autoimmune thyroiditis; diabetes mellitus; Crohn's disease;	
KW	multiple sclerosis; rheumatoid arthritis; ulcerative colitis;	
KW	cardiovascular disorder; wound healing; neurological disease.	
XX	Homo sapiens.	
OS	WO200055173-AI.	
NN	21-SEP-2000.	
PD	08-MAR-2000; 2000WO-US05881.	
XX	12-MAR-1999; 99US-0124270.	
PF	(HUMA-) HUMAN GENOME SCI INC.	
PP	Rosen CA, Ruben SM;	
PA	WIPI: 2000-611515/58.	
XX	N-PSDB; AAF21873.	
DR	New human breast and ovarian cancer associated gene sequences and the	
XX	polypeptides encoded by these genes, useful in the prevention,	
PT	treatment and diagnosis of cancer, immune disorders, cardiovascular	
PT	disorders and neurological diseases -	
XX	Claim 11; Page 1128-1132; 1299pp; English.	
PS	Sequences AAF21614 - AAF22031 represent DNA sequences encoding human	
CC	proteins AAB58711 - AAB59128. The DNA and protein sequences are	
CC	associated with breast and ovarian cancer. Included in the invention are	
CC	sequences AAF22032 - AAF22040 and AAB59129 which are used in the	
CC	isolation and characterisation of the DNA and protein sequences of the	
CC	invention. The breast and ovarian cancer associated DNA, protein, agonist	
CC	or antagonist sequences exhibit cytostatic; immunosuppressive;	
CC	nootropic; neuroprotective; antiviral; antiatherogenic; hepatotropic;	
CC	anti-diabetic; anti-inflammatory; anti-tumor; vulnerable; anticarcinoma;	
CC	antibacterial; antifungal; antiparasitic and cardiant activity. The	
CC	polynucleotide and protein sequences are used in the diagnosis of cancer,	
CC	particularly breast and ovarian cancer. The nucleic acid sequences,	
CC	proteins, agonists and antagonists may also be used in the diagnosis,	
CC	prevention and treatment of immune disorders e.g. Addison's disease,	
CC	allergies, autoimmune hemolytic anemia, autoimmune thyroditis,	
CC	diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid	
CC	arthritis and ulcerative colitis; cardiovascular disorders such as	
CC	myocardial ischemia; wound healing; neurological diseases such as	
CC	cerebral anoxia and epilepsy; and infectious diseases.	
XX	Sequence 940 AA;	
SQ	Query Match            24.0%; Score 83.5; DB 21; Length 940;	
	Best Local Similarity 42.9%; Pred. No. 1.6;	
	Matches 24; Conservative 9; Mismatches 16; Indels 7; Gaps 4;	
Qy	12 PETTIPSS-PPQWKLOP-GDDPLGNQSLLLETH-PLYQSEPAVYLTTPPKKK 64   :       :   : :   :   :   :   :   :   :   :   :   :   :   :   :	
Dd	525 PQITMPPSAPPRTQTTPGTGTPOLG---LKTNPPLIQEAKPYTKKPPSKKEE 576   :       :   : :   :   :   :   :   :   :   :   :   :   :   :   :	
RESULT 7		
AIDB0516	ABB60516 standard; Protein; 1190 AA.	
AC AC	ABB60516;	
XX XX	26-MAR-2002 (first entry)	
DT DT	Drosophila melanogaster polypeptide SEQ ID NO 8340.	
DE DE	Drosophila; developmental biology; cell signalling; insecticide;	
KW KW	pharmaceutical.	



XX OS Drosophila melanogaster.  
XX PN WO200171042-A2.  
XX PD 27-SEP-2001.  
XX PF 23-MAR-2001; 2001WO-US09231.  
XX PR 23-MAR-2000; 2000US-191637P.  
XX PR 11-JUL-2000; 2000US-0614150.  
XX PA (PEKE ) PE CORP NY.  
XX PI Venter JC, Adams M, Li PWD, Myers EW.  
XX DR WPI; 2001-656860/75.  
XX DR N-PSDB; ABL04619.  
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -  
XX PS Disclosure; SEQ ID NO 8340; 21pp + Sequence Listing; English.  
XX CC The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072).  
XX CC The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO\* at ftp.wipo.int/pub/published\_pct\_sequences.  
XX SQ Sequence 1190 AA;  
Query Match 22.1%; Score 77; DB 22; Length 1190;  
Best Local Similarity 35.5%; Pred. No. 11;  
Matches 27; Conservative 9; Mismatches 22; Indels 18; Gaps 5;  
Qy 5 RYQERPPETTTIPSSPPQKWL-----QPGD---DPLLGNSLLETH-----PLYQS 49  
Db 1060 RMEAKKFGERASQSANSPPQKLEPAENENFAKPADVTGDPDVA-KSQTETSAACTPLYDE 1118  
Qy 50 --EPAPVVIKTPPLKK 63  
Db 1119 LPPPAVQTPTTPPKKE 1134  
RESULT 8  
AAR42055  
ID AAR42055 standard; Protein; 280 AA.  
XX AC AAR42055;  
XX DT 03-MAY-1994 (first entry)  
XX DE Insecticidal/Fungicidal protein.  
XX KW Promoter; extensin; insect-resistance; fungicidal; insecticidal;  
XX KW beta-GUS.  
XX OS Nicotiana tabacum.  
XX PN JP05236964-A.  
XX PD 17-SEP-1993.  
XX PF 29-FEB-1992; 92JP-0078926.  
XX

PR 29-FEB-1992; 92JP-0078926.  
XX (NISB ) JAPAN TOBACCO INC.  
XX WPI; 1993-330584/42.  
XX DR N-PSDB; AAQ49903.  
XX PT Plant originated promoter which controls generation of extensin - used to produce noxious insect-resistant plant by connecting insecticidal and fungicidal protein gene, downstream of promoter  
XX PS Claim 1; Page 7; 11pp; Japanese.  
XX CC The activity of the promoter was confirmed by connecting it to the beta-GUS gene and determining the GUS activity of the resulting transformant. The sequence can be used to produce; noxious insect-resistant plants, by connecting an insecticidal and fungicidal protein downstream of the promoter.  
XX SQ Sequence 280 AA;  
Query Match 21.8%; Score 76; DB 14; Length 280;  
Best Local Similarity 40.0%; Pred. No. 2.7;  
Matches 24; Conservative 6; Mismatches 24; Indels 6; Gaps 4;  
Qy 6 YQERPPETTTIPSSPPQKWLQPGDPLLGNSLLETHPLYQS-EPAPVVIKTPPLKK 64  
Db 84 YKSPPPKPKYYP-HPVYKSP---PPKKPYSLPHT-PVYKSPPTPVYKSPPPPKK 138  
RESULT 9  
ABB91504  
ID ABB91504 standard; Protein; 708 AA.  
XX AC ABB91504;  
XX DT 31-MAY-2002 (first entry)  
XX DE Herbicidally active polypeptide SEQ ID NO 715.  
XX KW Herbicidal; plant; agriculture; herbicide.  
XX OS Arabidopsis thaliana.  
XX PN WO200210210-A2.  
XX PD 07-FEB-2002.  
XX PF 28-AUG-2001; 2001WO-EP09892.  
XX PR 28-AUG-2001; 2001WO-EP09892.  
XX PA (FARB ) BAYER AG.  
XX PI Tietjen K, Weidler M;  
XX WPI; 2002-269010/31.  
XX PT Identifying plant target proteins for herbicidally active compounds, comprising aligning and comparing nucleic acid or amino acid sequences from plant with nucleic acid or amino acid sequences from non-plant organisms -  
XX PS Claim 5; SEQ ID NO 715; 261pp + Sequence Listing; English.  
XX CC The invention relates to identifying target proteins (ABB90790-ABB94016) for herbicidally active compounds, comprising aligning and comparing nucleic acid or amino acid sequences from plant with nucleic acid or amino acid sequences from non-plant organisms using suitable search parameters, where plant sequences having an E-value greater by a factor of 3 than the E-value of most similar non-plant sequences are selected. The polypeptides or nucleic acids encoding them are useful for identifying modulators. The identified modulators are



Matches 22; Conservative 13; Mismatches 23; Indels 21; Gaps 3;

QY 4 RRYQERPPETTTI-----PPSSPPQWKLPQGDD----PLIGNQSILLETHTPL----- 46  
:: : :: : :: : ||| : :: : ||| : :: : ||| : :: : |||

Db 63 QQQQQOQQPQATLTAPQPQVPTTQQVPSPSQQAQTLLVVQPMQSSPLSLPPDAAPK 122

QY 47 ----YQSEPAVPVIKTPL 61  
||:| | ||| |

Db 123 PPIPIQSKPPVAPIKPPQL 141

RESULT 12

ABBI18119

ID ABB18119 standard; Protein; 223 AA.

XX ABB18119;

AC XX

DT 23-JAN-2002 (first entry)

XX

XX Protein #118 encoded by probe for measuring heart cell gene expression.

XX Human; gene expression; heart; microarray; vascular system;  
XX cardiovascular disease; hypertension; cardiac arrhythmia;  
XX congenital heart disease.

OS Homo sapiens.

XX WO200157274-A2.

PN

XX

PD 09-AUG-2001.

XX

PF 30-JAN-2001; 2001WO-US00666.

XX

PR 04-FEB-2000; 2000US-0180312.  
26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PA

XX Penn SG, Hanzel DK, Chen W, Rank DR;

PI

XX WPI; 2001-488899/53.

DR

XX Single exon nucleic acid probes for analyzing gene expression in human hearts -

PT

XX Claim 15; SEQ ID No 19889; 530pp; English.

PS

XX The present invention relates to single exon nucleic acid probes for measuring human gene expression in a sample derived from human heart (see ABA21533-ABA41305). The present sequence is a protein encoded by one such probe. The probes may be used for predicting, measuring and displaying gene expression in samples derived from the human heart via microarrays. By measuring gene expression, the probes are useful for predicting, diagnosing, grading, staging, monitoring and prognosing diseases of the human heart and vascular system e.g. cardiovascular disease, hypertension, cardiac arrhythmias and congenital heart disease.

CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

CC

XX Sequence 223 AA;

SQ

Query Match 21.4%; Score 74.5; DB 22; Length 223;  
Best Local Similarity 27.8%; Pred. No. 3;  
Matches 22; Conservative 13; Mismatches 23; Indels 21; Gaps 3;

QY 4 RRYQERPPETTTI-----PPSSPPQWKLPQGDD----PLIGNQSILLETHTPL----- 46  
:: : :: : :: : ||| : :: : ||| : :: : ||| : :: : |||

```
ID AM65828 standard; Protein; 223 AA.
XX AC AM65828;
XX DT 06-NOV-2001 (first entry)
XX DE Human bone marrow expressed probe encoded protein SEQ ID NO: 26134.
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;
XX KW microarray; cancer; leukaemia; lymphoma; myeloma.
XX OS Homo sapiens.
XX PN WO200157276-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00668.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-488900/53.
XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human bone marrow.
XX PS Example 4; SEQ ID NO: 26134; 658pp + Sequence Listing; English.
XX CC The present invention provides a number of single exon nucleic acid
XX CC probes which are derived from genomic sequences expressed in the human
XX CC bone marrow. They can be used to measure gene expression in bone marrow
XX CC samples, which may enable the improved diagnosis and treatment of cancers
XX CC such as lymphoma, leukaemia and myeloma. The present sequence is a
XX CC protein encoded by one of the probes of the invention.
XX SQ Sequence 223 AA;
Query Match 21.4%; Score 74.5; DB 22; Length 223;
Best Local Similarity 27.8%; Pred. No. 3;
Matches 22; Conservative 13; Mismatches 23; Indels 21; Gaps 3;
QY 4 RRYQERPPETTTI----PPSSPPQWKLPQGDGDD----PLLGNSLLETHPL----- 46
DB 63 QQQQQQQQQTATLTAPQPPVPTQQVPPSQSQQAOTLVVQPMQLSSLSLPPDAPK 122
QY 47 ----YQSEPAVPVTKTTPPL 61
DB 123 PPIPIQSKPPVAPIKPPQL 141
RESULT 15
AAM13690
ID AAM13690 standard; Protein; 223 AA.
XX AC AAM13690;
XX DT 12-OCT-2001 (first entry)
XX DE Peptide #124 encoded by probe for measuring cervical gene expression.
XX KW Probe; human; microarray; gene expression; cervical epithelial cell;
XX KW cervical cancer.
```

```
OS Homo sapiens.
XX PN WO200157278-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00670.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-488901/53.
XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human cervical epithelial cells -
XX PS Claim 27; SEQ ID No 18516; 487pp; English.
XX CC The present invention relates to human single exon nucleic acid probes
XX CC (SENPs: see AAI10068-AAI128459). The present sequence is a peptide encoded
XX CC by one such probe. The SENPs are derived from human Hela cells. The SENPs
XX CC can be used to produce a single exon microarray, which can be used for
XX CC measuring human gene expression in a sample derived from human cervical
XX CC epithelial cells. By measuring gene expression, the probes are therefore
XX CC useful in grading and/or staging of diseases of the cervix, notably
XX CC cervical cancer.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 223 AA;
Query Match 21.4%; Score 74.5; DB 22; Length 223;
Best Local Similarity 27.8%; Pred. No. 3;
Matches 22; Conservative 13; Mismatches 23; Indels 21; Gaps 3;
QY 4 RRYQERPPETTTI----PPSSPPQWKLPQGDGDD----PLLGNSLLETHPL----- 46
DB 63 QQQQQQQQQTATLTAPQPPVPTQQVPPSQSQQAOTLVVQPMQLSSLSLPPDAPK 122
QY 47 ----YQSEPAVPVTKTTPPL 61
DB 123 PPIPIQSKPPVAPIKPPQL 141
Search completed: January 2, 2003, 13:44:27
Job time : 24.9677 secs
```

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	348	100.0	161	2	US-08-683-262B-34	Sequence 34, Appl
2	348	100.0	161	4	US-09-361-707-34	Sequence 34, Appl
3	104	29.9	48	2	US-08-683-262B-68	Sequence 68, Appl
4	104	29.9	48	4	US-09-361-707-68	Sequence 68, Appl
5	83.5	24.0	906	2	US-08-609-230A-9	Sequence 9, Appl
6	83.5	24.0	907	3	US-08-990-140-4	Sequence 4, Appl
7	83.5	24.0	907	4	US-09-546-238-4	Sequence 4, Appl
8	83.5	24.0	940	4	US-08-810-712-7	Sequence 7, Appl
9	74.5	21.4	1004	4	US-08-916-332-2	Sequence 2, Appl
10	73	21.0	109	4	US-09-187-859-15	Sequence 15, Appl
11	73	21.0	794	1	US-08-188-228-60	Sequence 60, Appl
12	73	21.0	794	1	US-08-333-643-54	Sequence 54, Appl
13	73	21.0	794	1	US-08-333-638-60	Sequence 60, Appl
14	70	20.1	109	4	US-09-187-859-11	Sequence 11, Appl
15	69	19.8	652	6	5202336-13	Patent No. 5202336
16	69	19.8	1158	4	US-09-060-482-2	Sequence 2, Appl
17	68.5	19.7	214	1	US-08-217-327-4	Sequence 4, Appl
18	68	19.5	605	2	US-08-687-956A-1	Sequence 1, Appl
19	67.5	19.4	210	4	US-09-071-035-232	Sequence 232, App
20	67.5	19.4	256	4	US-09-071-035-230	Sequence 230, App
21	67.5	19.4	525	4	US-08-764-870-7	Sequence 7, Appl
22	67.5	19.4	525	4	US-08-980-115-7	Sequence 7, Appl
23	67	19.3	631	4	US-09-147-119-7	Sequence 7, Appl
24	67	19.3	744	6	5202336-25	Patent No. 5202336
25	67	19.3	1257	1	US-08-340-428B-49	Sequence 49, Appl
26	65.5	18.8	281	4	US-09-071-035-224	Sequence 224, App
27	65.5	18.8	282	4	US-09-071-035-222	Sequence 222, App

Db 158 LKKK 161

## RESULT 2

US-09-361-707-34

; Sequence 34, Application US/09361707

; Patent No. 6258937

; GENERAL INFORMATION:

; APPLICANT: Tong, Shuping

; Li, Jisu

; Wands, Jack R.

; TITLE OF INVENTION: HEPA DNA VIRUS RECEPTOR

; NUMBER OF SEQUENCES: 117

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish &amp; Richardson P.C.

; STREET: 225 Franklin Street

; CITY: Boston

; STATE: MA

; COUNTRY: USA

; ZIP: 02110-2804

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/361.707

; FILING DATE: 27-Jul-1999

; PRIORITY APPLICATION NUMBER: 08/683,262

; FILING DATE: 18-Jul-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Creason, Gary L.

; REGISTRATION NUMBER: 34,310

; REFERENCE/DOCKET NUMBER: 00786/287003

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (617) 542-5070

; TELEFAX: (617) 542-8906

; INFORMATION FOR SEQ ID NO: 34:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 161 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; FRAGMENT TYPE: internal

; SEQUENCE DESCRIPTION: SEQ ID NO: 34:

; US-09-361-707-34

; Query Match 100.0%; Score 348; DB 4; Length 161;

; Best Local Similarity 100.0%; Pred. No. 4.8e-34;

; Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

; QY 1 EAFRRYQERPPETTTIPSSPPQKLPQDDPLLGNSLLETHPLYOSEPAVPVKTIPP 60

; Db 98 EAFRRYQERPPETTTIPSSPPQKLPQDDPLLGNSLLETHPLYOSEPAVPVKTIPP 157

; QY 61 LKKK 64

; Db 158 LKKK 161

; RESULT 3

US-08-683-262B-68

; Sequence 68, Application US/08683262B

; Patent No. 5929220

; GENERAL INFORMATION:

; APPLICANT: Shuping Tong et al.

; TITLE OF INVENTION: HEPADNAVIRUS RECEPTOR

; NUMBER OF SEQUENCES: 75

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish &amp; Richardson P.C.

; STREET: 225 Franklin Street

; CITY: Boston

; STATE: MA

; COUNTRY: USA

; ZIP: 02110-2804

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/361.707

; FILING DATE: 27-Jul-1999

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/683,262

; FILING DATE: 18-Jul-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Creason, Gary L.

; REGISTRATION NUMBER: 34,310

; REFERENCE/DOCKET NUMBER: 00786/287003

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (617) 542-5070

; TELEFAX: (617) 542-8906

; INFORMATION FOR SEQ ID NO: 68:

; SEQUENCE CHARACTERISTICS:

; STATE: Massachusetts

; COUNTRY: U.S.A.

; ZIP: 02110-2804

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/683,262B

; FILING DATE: 18-Jul-1996

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Fraser, Janis K.

; REGISTRATION NUMBER: 31,819

; REFERENCE/DOCKET NUMBER: 00786/287002

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (617) 542-5070

; TELEFAX: (617) 542-8906

; TELEX: 200154

; INFORMATION FOR SEQ ID NO: 68:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 48 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; US-08-683-262B-68

; Query Match 29.9%; Score 104; DB 2; Length 48;

; Best Local Similarity 100.0%; Pred. No. 9.4e-06;

; Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

; QY 1 EAFRRYQERPPETTTIPP 19

; Db 30 EAFRRYQERPPETTTIPP 48

; RESULT 4

US-09-361-707-68

; Sequence 68, Application US/09361707

; Patent No. 6258937

; GENERAL INFORMATION:

; APPLICANT: Tong, Shuping

; Li, Jisu

; Wands, Jack R.

; TITLE OF INVENTION: HEPA DNA VIRUS RECEPTOR

; NUMBER OF SEQUENCES: 117

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish &amp; Richardson P.C.

; STREET: 225 Franklin Street

; CITY: Boston

; STATE: MA

; COUNTRY: USA

; ZIP: 02110-2804

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/361.707

; FILING DATE: 27-Jul-1999

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/683,262

; FILING DATE: 18-Jul-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Creason, Gary L.

; REGISTRATION NUMBER: 34,310

; REFERENCE/DOCKET NUMBER: 00786/287003

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (617) 542-5070

; TELEFAX: (617) 542-8906

; INFORMATION FOR SEQ ID NO: 68:

; SEQUENCE CHARACTERISTICS:

LENGTH: 48 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 68:  
US-09-361-707-68

Query Match 29.9%; Score 104; DB 4; Length 48;  
Best Local Similarity 100.0%; Pred. No. 9.4e-06;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EAFRRYQERPPETTTIPP 19  
| | | | | | | | | | | | | | | | | | | | | |  
Db 30 EAFRRYQERPPETTTIPP 48

RESULT 5  
US-08-609-230A-9  
; Sequence 9, Application US/08609230A  
; Patent No. 5866333  
; GENERAL INFORMATION:  
; APPLICANT: Innerarity, Thomas L.  
; APPLICANT: Qian, Xiaobing  
; APPLICANT: Yamanaka, Shinya  
; TITLE OF INVENTION: Screening Methods to Detect mRNA Targets  
; TITLE OF INVENTION: Of Editing Enzymes  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, 8th Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/609,230A  
; FILING DATE: 01-MAR-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fitts, Renee A.  
; REGISTRATION NUMBER: 35,136  
; REFERENCE/DOCKET NUMBER: 02307U-06810005  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-326-2400  
; TELEFAX: 650-326-2422  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 906 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-609-230A-9

Query Match 24.0%; Score 83.5; DB 2; Length 906;  
Best Local Similarity 42.9%; Pred. No. 0.09;  
Matches 24; Conservative 9; Mismatches 16; Indels 7; Gaps 4;

QY 12 PETTITPSS-PPQWKLOP-GDDPLLGNSLLETH-PLYQSEPAVPVTKTPPLKKK 64  
| | | | | | | | | | | | | | | | | | | | | |  
Db 491 PQTMIIPPSAQPPRTQTPPLGQTPLG----LKNPPLIQEKPAKTSKKPPPSKEE 542

RESULT 6  
US-08-990-140-4  
; Sequence 4, Application US/08990140A  
; Patent No. 6093795  
; GENERAL INFORMATION:  
; APPLICANT: Olsen, Henrik S.

; APPLICANT: Ruben, Steven M.  
; APPLICANT: Sonenberg, Nahum  
; APPLICANT: Methot, Nathalie  
; APPLICANT: Rom, Eran  
; TITLE OF INVENTION: Human Prt1-like Subunit Protein (hPrt1) and Human  
; TITLE OF INVENTION: eif4Gf-like Protein (p97) Genes  
; FILE REFERENCE: 1488.0700001  
; CURRENT APPLICATION NUMBER: US/08/990,140A  
; CURRENT FILING DATE: 1997-12-12  
; EARLIER APPLICATION NUMBER: US 60/033,151  
; EARLIER FILING DATE: 1996-12-13  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 907  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-08-990-140-4

Query Match 24.0%; Score 83.5; DB 3; Length 907;  
Best Local Similarity 42.9%; Pred. No. 0.09;  
Matches 24; Conservative 9; Mismatches 16; Indels 7; Gaps 4;

QY 12 PETTITPSS-PPQWKLOP-GDDPLLGNSLLETH-PLYQSEPAVPVTKTPPLKKK 64  
| | | | | | | | | | | | | | | | | | | | | |  
Db 492 PQTMIIPPSAQPPRTQTPPLGQTPLG----LKNPPLIQEKPAKTSKKPPPSKEE 543

RESULT 7  
US-09-546-238-4  
; Sequence 4, Application US/09546238  
; Patent No. 6316225  
; GENERAL INFORMATION:  
; APPLICANT: Olsen, Henrik S.  
; APPLICANT: Ruben, Steven M.  
; APPLICANT: Sonenberg, Nahum  
; APPLICANT: Methot, Nathalie  
; APPLICANT: Rom, Eran  
; TITLE OF INVENTION: Human Prt1-like Subunit Protein (hPrt1) Polynucleotides  
; FILE REFERENCE: 1488.0700002  
; CURRENT APPLICATION NUMBER: US/09/546,238  
; CURRENT FILING DATE: 2000-04-10  
; PRIOR APPLICATION NUMBER: US 60/033,151  
; PRIOR FILING DATE: 1996-12-13  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 907  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-546-238-4

Query Match 24.0%; Score 83.5; DB 4; Length 907;  
Best Local Similarity 42.9%; Pred. No. 0.09;  
Matches 24; Conservative 9; Mismatches 16; Indels 7; Gaps 4;

QY 12 PETTITPSS-PPQWKLOP-GDDPLLGNSLLETH-PLYQSEPAVPVTKTPPLKKK 64  
| | | | | | | | | | | | | | | | | | | | | |  
Db 492 PQTMIIPPSAQPPRTQTPPLGQTPLG----LKNPPLIQEKPAKTSKKPPPSKEE 543

RESULT 8  
US-08-810-712-7  
; Sequence 7, Application US/08810712G  
; Patent No. 6160106  
; GENERAL INFORMATION:  
; APPLICANT: Yeda Research and Development Co. LTD  
; TITLE OF INVENTION: Tumor Suppressor Genes, Proteins Encoded Thereby and  
; TITLE OF INVENTION: Use of said Genes and Proteins  
; FILE REFERENCE: sequencelist  
; CURRENT APPLICATION NUMBER: US/08/810,712G  
; CURRENT FILING DATE: 1997-03-03





```
; MOLECULE TYPE: protein
; US-08-188-228-60

Query Match      21.0%; Score 73; DB 1; Length 794;
Best Local Similarity 33.3%; Pred. No. 1.4;
Matches 19; Conservative 6; Mismatches 22; Indels 10; Gaps 2;

QY 12 PETTTIPSSPP-----QWKLPQGDGDPPLGNQ-----SLLETHPLYQSEPAVPVKT 58
Db 164 PYVATVPMSPGVAYVLQVKATDADPTYGNSARVVYSILQGPYFSIDPKTGVIRT 220

RESULT 12
US-08-332-643-54
; Sequence 54, Application US/08332643
; Patent No. 5639634
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Bicknell
; STREET: Two First National Plaza, 20 South Clark
; STREET: Street
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/332.643
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872.643
; FILING DATE: 17 APR 1992
; APPLICATION NUMBER: US/08/049.460
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5646250and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 31340
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 60:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 794 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-332-638-60

Query Match      21.0%; Score 73; DB 1; Length 794;
Best Local Similarity 33.3%; Pred. No. 1.4;
Matches 19; Conservative 6; Mismatches 22; Indels 10; Gaps 2;

QY 12 PETTTIPSSPP-----QWKLPQGDGDPPLGNQ-----SLLETHPLYQSEPAVPVKT 58
Db 164 PYVATVPMSPGVAYVLQVKATDADPTYGNSARVVYSILQGPYFSIDPKTGVIRT 220

RESULT 14
US-09-187-859-11
; Sequence 11, Application US/09187859A
; Patent No. 6358920
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING NONCLASSICAL
; FILE OF INVENTION: CADHERIN-MEDIATED FUNCTIONS
; FILE REFERENCE: 100086.407C1
; CURRENT APPLICATION NUMBER: US/09/187,859A
; CURRENT FILING DATE: 1998-11-06
; NUMBER OF SEQ ID NOS: 4052
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 11
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Gallus gallus
; US-09-187-859-11

Query Match      20.1%; Score 70; DB 4; Length 109;
Best Local Similarity 33.3%; Pred. No. 0.27;
Matches 19; Conservative 5; Mismatches 23; Indels 10; Gaps 2;

QY 12 PETTTIPSSP-----PQWKLQPGDGPPLGNQ-----SLLETHPLYQSEPAVPVKT 58
```

Db 4 PYTAGVEMSPVGTSVQVATDADDPYGNARSARVYVSILOQPYFSVEPKTGIIKT 60

RESULT 15

5202236-13

; Patent No. 5202236

; APPLICANT: MAUGH, KATHY J.; ANDERSON, DAVID M.; STRAUSBERG,

; SUSAN L.; MCCANDLISS, RUSS; WEI, TENA; FILPULA, DAVID

; TITLE OF INVENTION: METHOD OF PRODUCING BIOADHESIVE

; PROTEIN

; NUMBER OF SEQUENCES: 39

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/528,762

; FILING DATE: 25-MAY-1990

; APPLICATION NUMBER: 82,456

; FILING DATE: 07-AUG-1987

; APPLICATION NUMBER: 933,945

; FILING DATE: 24-NOV-1986

; APPLICATION NUMBER: 650,128

; FILING DATE: 13-SEP-1984

; SEQ ID NO:13:

; LENGTH: 652

5202236-13

Query Match

Best Local Similarity 19.8%; Score 69; DB 6; Length 652;

Matches 16; Conservative 8; Mismatches 25; Indels 4; Gaps 1;

QY 12 PETTTIPSSPPQWKLOPGDDPLLGNQSLLETHPLYQSEPAVPVIKTPPLKKK 64

Db 494 PTYKAKPSYPTYKAKPSYPTYKAKTYKAKTYKAKPSYP----PTYKAK 542

Search completed: January 2, 2003, 13:47:33

Job time : 9.51613 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2003, 13:45:43 ; Search time 5.41935 Seconds  
(without alignments)  
223.808 Million cell updates/sec

Title: US-09-818-066-34\_COPY\_98\_161  
Perfect score: 348  
Sequence: 1 EAFRRQEEPPPTTIPPS.....PLTQSEPAVPVITPPLKKK 64

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 117078 seqs, 18951520 residues

Total number of hits satisfying chosen parameters: 117078

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCIT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/PCIT\_PUBCOMB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	348	100.0	161	10	US-09-818-066-34
2	104	29.9	48	10	US-09-818-066-68
3	83.5	24.0	907	10	US-09-954-043-4
4	74.5	21.4	223	10	US-09-864-761-33417
5	72.5	20.8	886	10	US-09-801-368-180
6	72.5	20.8	2665	10	US-09-864-761-34248
7	70.5	20.3	920	10	US-09-983-531A-4
8	67.5	19.4	758	10	US-09-801-368-224
9	67	19.3	780	10	US-09-770-689A-5
10	65.5	18.8	726	10	US-09-770-689A-4
11	65.5	18.8	803	10	US-09-770-689A-2
12	65.5	18.8	881	10	US-09-816-860A-2
13	64	18.4	281	8	US-08-971-317A-6
14	64	18.4	281	10	US-09-802-669-25
15	64	18.4	281	10	US-09-193-663-6
16	64	18.4	281	10	US-09-027-287-6
17	64	18.4	281	10	US-09-252-656B-6
18	64	18.4	281	10	US-09-929-493-6
19	64	18.4	281	10	US-09-927-110-1

20	64	18.4	281	12	US-10-012-452-13
21	63.5	18.2	359	12	US-10-029-654-12
22	63.5	18.2	1056	9	US-10-161-510-10
23	63.5	18.2	1317	10	US-09-963-896-7
24	63	18.1	592	9	US-09-738-626-4709
25	62.5	18.0	1337	10	US-09-803-126-1
26	61.5	17.7	273	10	US-09-764-864-1421
27	61.5	17.7	316	10	US-09-764-864-1036
28	61.5	17.7	682	10	US-09-920-552-13
29	61.5	17.7	731	9	US-10-086-464-17
30	61.5	17.7	1814	10	US-09-920-552-103
31	61	17.5	115	10	US-09-764-869-634
32	61	17.5	236	10	US-09-731-872-412
33	61	17.5	402	10	US-09-933-561-16
34	61	17.5	404	10	US-09-764-864-1368
35	61	17.5	772	9	US-09-978-295A-264
36	61	17.5	772	9	US-09-978-697-264
37	61	17.5	772	9	US-09-978-192A-264
38	61	17.5	772	9	US-10-086-500-121
39	61	17.5	772	9	US-09-999-832A-264
40	61	17.5	772	12	US-10-052-586-106
41	61	17.5	819	10	US-09-825-144-14
42	61	17.5	941	12	US-10-124-557-14
43	61	17.5	1022	12	US-10-124-557-84
44	61	17.5	1038	12	US-10-124-557-74
45	61	17.5	1049	12	US-10-124-557-58

ALIGNMENTS

RESULT 1

US-09-818-066-34

Sequence 34, Application US/09818066

Patent No. US20020032307A1

GENERAL INFORMATION:

APPLICANT: Shuping Tong et al.

TITLE OF INVENTION: HEPADNAVIRUS RECEPTOR

NUMBER OF SEQUENCES: 75

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson P.C.

STREET: 225 Franklin Street

CITY: Boston

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/818,066

FILING DATE: 27-Mar-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/683,262

FILING DATE: 18-JUL-1996

ATTORNEY/AGENT INFORMATION:

NAME: Fraser, Janis K.

REGISTRATION NUMBER: 31,819

REFERENCE/DOCKET NUMBER: 00786/287002

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 542-5070

TELEFAX: (617) 542-8906

TELEX: 200154

INFORMATION FOR SEQ ID NO: 34:

SEQUENCE CHARACTERISTICS:

LENGTH: 161 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 34:



;; PRIOR APPLICATION NUMBER: PCT/US01/00661  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00670  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: US 60/234,687  
;; PRIOR FILING DATE: 2000-09-21  
;; PRIOR APPLICATION NUMBER: US 09/608,408  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: US 09/774,203  
;; PRIOR FILING DATE: 2001-01-29  
;; NUMBER OF SEQ ID NOS: 49117  
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
;; SEQ ID NO 33417  
;; LENGTH: 223  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; OTHER INFORMATION: MAP TO AC006581.16  
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 13  
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 41  
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 23  
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 25  
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 28  
;; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 46  
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 17  
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 71  
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 15  
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 29  
;; OTHER INFORMATION: EST\_HUMAN HIT: AW959289.1, EVALUE 9.00e-15  
US-09-864-761-33417

Query Match 21.4%; Score 74.5; DB 10; Length 223;  
Best Local Similarity 27.8%; Pred. No. 0.63;  
Matches 22; Conservative 13; Mismatches 23; Indels 21; Gaps 3;

QY 4 RRYOEPRPPTTTI---PSSPPQWKLPQD---PILGNQSLLETHPL----- 46  
DB 63 QQQQQQQQQPQATTTATAPQVPTQVPPSQSQQAQTLVQVPLQSSPLSLPPDAAPK 122  
QY 47 ----YQSEPAVPVTKTPL 61  
DB 123 PPIPIQSKPPVAPIKPQL 141

RESULT 5  
US-09-801-368-180  
;; Sequence 180, Application US/09801368  
;; Patent No. US20020128250A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Busby, Robert  
;; APPLICANT: Cali, Brian  
;; APPLICANT: Hecht, Peter  
;; APPLICANT: Holtzman, Doug  
;; APPLICANT: Madden, Kevin  
;; APPLICANT: Maxon, Mary  
;; APPLICANT: Milne, Todd  
;; APPLICANT: No. US20020128250A1man, Thea  
;; APPLICANT: Royer, John  
;; APPLICANT: Salama, Sofie  
;; APPLICANT: Sherman, Amir  
;; APPLICANT: Silva, Jeff  
;; APPLICANT: Summers, Eric  
;; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi  
;; FILE REFERENCE: 109272.147  
;; CURRENT APPLICATION NUMBER: US/09/801,368  
;; CURRENT FILING DATE: 2001-03-07  
;; PRIOR APPLICATION NUMBER: US 09/487,558  
;; PRIOR FILING DATE: 2000-01-19  
;; PRIOR APPLICATION NUMBER: US 60/160,587  
;; PRIOR FILING DATE: 1999-10-20  
;; NUMBER OF SEQ ID NOS: 440  
;; SOFTWARE: PatentIn version 3.0  
;; SEQ ID NO 180

;; LENGTH: 886  
;; TYPE: PRT  
;; ORGANISM: Saccharomyces cerevisiae  
US-09-801-368-180

Query Match 20.8%; Score 72.5; DB 10; Length 886;  
Best Local Similarity 39.1%; Pred. No. 5.1;  
Matches 18; Conservative 5; Mismatches 20; Indels 3; Gaps 1;

QY 18 PSSPPQWKLPQDGPL---LGNQSLLETHPLYQSEPAVPVTKTTP 60  
DB 809 PANOPPLSIYQMQLPATQANSLLLETPYVQSNPVTYTTIKESP 854

## RESULT 6

US-09-864-761-34248  
;; Sequence 34248, Application US/09864761  
;; Patent No. US20020048763A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Penn, Sharron G.  
;; APPLICANT: Rank, David R.  
;; APPLICANT: Hanzel, David K.  
;; APPLICANT: Chen, Wensheng  
;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
;; FILE REFERENCE: Aeonica-X-1  
;; CURRENT APPLICATION NUMBER: US/09/864,761  
;; CURRENT FILING DATE: 2001-05-23  
;; PRIOR APPLICATION NUMBER: US 60/180,312  
;; PRIOR FILING DATE: 2000-02-04  
;; PRIOR APPLICATION NUMBER: US 60/207,456  
;; PRIOR FILING DATE: 2000-05-26  
;; PRIOR APPLICATION NUMBER: US 09/632,366  
;; PRIOR FILING DATE: 2000-08-03  
;; PRIOR APPLICATION NUMBER: GB 24263.6  
;; PRIOR FILING DATE: 2000-10-04  
;; PRIOR APPLICATION NUMBER: US 60/236,359  
;; PRIOR FILING DATE: 2000-09-27  
;; PRIOR APPLICATION NUMBER: PCT/US01/00666  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00667  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00664  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00669  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00665  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00668  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00663  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00662  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00661  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00670  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: US 60/234,687  
;; PRIOR APPLICATION NUMBER: US 09/608,408  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: US 09/774,203  
;; PRIOR FILING DATE: 2001-01-29  
;; NUMBER OF SEQ ID NOS: 49117  
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
;; SEQ ID NO 34248  
;; LENGTH: 2665  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; OTHER INFORMATION: MAP TO AL034555.2  
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 10

; APPLICANT: Busby, Robert

US-05-770-085A-4  
: Sequence 4. Application US/09770689A

```

; Patent No. US20020115171A1
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN RAS-LIKE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING THESE HUMAN RAS-LIKE
; TITLE OF INVENTION: PROTEINS, AND USES THEREOF
; FILE REFERENCE: CL001079
; CURRENT APPLICATION NUMBER: US/09/770,689A
; CURRENT FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 726
; TYPE: PRT
; ORGANISM: HUMAN
US-09-770-689A-4

Query Match      18.8%; Score 65.5; DB 10; Length 726;
Best Local Similarity 31.2%; Pred. No. 21;
Matches 20; Conservative 6; Mismatches 23; Indels 15; Gaps 3;

Qy 11 PPETTTPSSPPQWKLOPG-----DDPLLGNSLLETH-----PLQSEPAVPVI 56
Db 599 PPQTPT-PPSTPPLGKQNPSPAPQTLAGGNPETAQPHAGTLPRPRPVKPRNRPVSVP 657

Qy 57 KTHP 60
Db 658 PQPP 661

RESULT 11
US-09-770-689A-2
; Sequence 2, Application US/09770689A
; Patent No. US20020115171A1
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN RAS-LIKE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING THESE HUMAN RAS-LIKE
; TITLE OF INVENTION: PROTEINS, AND USES THEREOF
; FILE REFERENCE: CL001079
; CURRENT APPLICATION NUMBER: US/09/770,689A
; CURRENT FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 803
; TYPE: PRT
; ORGANISM: HUMAN
US-09-770-689A-2

Query Match      18.8%; Score 65.5; DB 10; Length 803;
Best Local Similarity 31.2%; Pred. No. 24;
Matches 20; Conservative 6; Mismatches 23; Indels 15; Gaps 3;

Qy 11 PPETTTPSSPPQWKLOPG-----DDPLLGNSLLETH-----PLQSEPAVPVI 56
Db 676 PPQTPT-PPSTPPLGKQNPSPAPQTLAGGNPETAQPHAGTLPRPRPVKPRNRPVSVP 734

Qy 57 KTHP 60
Db 735 PQPP 738

RESULT 12
US-09-816-860A-2
; Sequence 2, Application US/09816860A
; Patent No. US20020081651A1
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel
; TITLE OF INVENTION: 26649, A No. US20020081651A1el Human GTPase Activating Molecule
; TITLE OF INVENTION: Uses Therefor
; FILE REFERENCE: MNI-133
; CURRENT APPLICATION NUMBER: US/09/816,860A

```

```

; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: 60/191,859
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 881
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-816-860A-2

Query Match      18.8%; Score 65.5; DB 10; Length 881;
Best Local Similarity 31.2%; Pred. No. 27;
Matches 20; Conservative 6; Mismatches 23; Indels 15; Gaps 3;

Qy 11 PPETTTPSSPPQWKLOPG-----DDPLLGNSLLETH-----PLQSEPAVPVI 56
Db 754 PPQTPT-PPSTPPLGKQNPSPAPQTLAGGNPETAQPHAGTLPRPRPVKPRNRPVSVP 812

Qy 57 KTHP 60
Db 813 PQPP 816

RESULT 13
US-08-971-317A-6
; Sequence 6, Application US/08971317A
; Patent No. US20010010925A1
; GENERAL INFORMATION:
; APPLICANT: Wiley, Steven R.
; TITLE OF INVENTION: TNF-DELTA LIGAND AND USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION NUMBER: US/08/971,317A
; FILING DATE: 17-NOV-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Goller, Mimi C
; REGISTRATION NUMBER: 39,046
; REFERENCE/DOCKET NUMBER: 6255.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (847) 935-7550
; TELEFAX: (847) 938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 281 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. US20010010925A1e
US-08-971-317A-6

Query Match      18.4%; Score 64; DB 8; Length 281;
Best Local Similarity 31.1%; Pred. No. 10;
Matches 19; Conservative 4; Mismatches 14; Indels 24; Gaps 3;

```

QY 4 RRYQERPPETTIPSSPPQWKLOPGDDPLLGNSLLETHPLYQSEPAVPVTKTPLKK 63  
 || : ||| || || | : ||||  
 Db 38 RRPQRRPP-----PLPPPPPP-----PPLPPLPLPLKK 73  
 QY 64 K 64  
 Db 74 R 74

Db 38 RRPQRRPP-----PLPPPPPP-----PPLPPLPLPLKK 73  
 QY 64 K 64  
 Db 74 R 74

Search completed: January 2, 2003, 13:55:35  
 Job time : 6.41935 secs

RESULT 14  
 US-09-802-669-25  
 ; Sequence 25, Application US/09802669  
 ; Patent No. US20020004490A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dean, Nicholas M.  
 ; APPLICANT: Marcusson, Eric G.  
 ; APPLICANT: Wyatt, Jacqueline  
 ; APPLICANT: Zhang, Hong  
 ; TITLE OF INVENTION: Antisense Compound Modulation of Fas Mediated Signaling  
 ; FILE REFERENCE: ISPH-545  
 ; CURRENT APPLICATION NUMBER: US/09/802,669  
 ; CURRENT FILING DATE: 2001-03-09  
 ; PRIOR APPLICATION NUMBER: US 09/665,615  
 ; PRIOR FILING DATE: 2000-09-18  
 ; PRIOR APPLICATION NUMBER: US 09/290,640  
 ; PRIOR FILING DATE: 1999-04-12  
 ; NUMBER OF SEQ ID NOS: 180  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 25  
 ; LENGTH: 281  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-09-802-669-25

Query Match 18.4%; Score 64; DB 10; Length 281;  
 Best Local Similarity 31.1%; Pred. No. 10;  
 Matches 19; Conservative 4; Mismatches 14; Indels 24; Gaps 3;

QY 4 RRYQERPPETTIPSSPPQWKLOPGDDPLLGNSLLETHPLYQSEPAVPVTKTPLKK 63  
 || : ||| || || | : ||||  
 Db 38 RRPQRRPP-----PLPPPPPP-----PPLPPLPLPLKK 73  
 QY 64 K 64  
 Db 74 R 74

RESULT 15  
 US-09-193-663-6  
 ; Sequence 6, Application US/09193663  
 ; Patent No. US20020055624A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wiley, Steven R.  
 ; TITLE OF INVENTION: TNF-DELTA LIGAND AND USES THEREOF  
 ; FILE REFERENCE: 6255.US.02  
 ; CURRENT APPLICATION NUMBER: US/09/193,663  
 ; CURRENT FILING DATE: 1998-11-17  
 ; EARLIER APPLICATION NUMBER: 60/065,916  
 ; EARLIER FILING DATE: 1997-11-17  
 ; NUMBER OF SEQ ID NOS: 10  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 6  
 ; LENGTH: 281  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-09-193-663-6

Query Match 18.4%; Score 64; DB 10; Length 281;  
 Best Local Similarity 31.1%; Pred. No. 10;  
 Matches 19; Conservative 4; Mismatches 14; Indels 24; Gaps 3;

QY 4 RRYQERPPETTIPSSPPQWKLOPGDDPLLGNSLLETHPLYQSEPAVPVTKTPLKK 63  
 || : ||| || || | : ||||



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2003, 13:41:48 ; Search time 9.54839 Seconds  
(without alignments)  
644.360 Million cell updates/sec

Title: US-09-818-066-34\_COPY\_98\_161  
Perfect score: 348  
Sequence: 1 EAFRRYQERPEPTTTPPS.....PLYQSEPAVPVTKPLKKK 64

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR.73.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	348	100.0	364	1 SAVLWE	large surface anti
2	303.5	87.2	365	1 SAVLWE	large surface anti
3	302.5	86.9	366	1 SAVLWD	large surface anti
4	297.5	85.5	366	1 SAVLBD	large surface anti
5	169	48.6	138	2 JC1095	pre-S protein - du
6	136.5	39.2	335	1 SAVLHH	large surface anti
7	80.5	23.1	328	2 AB0810	cell division prot
8	76	21.8	224	2 T03238	extensin (clone Ex
9	76	21.8	280	2 T03236	extensin precursor
10	76	21.8	443	2 T05540	hypothetical prote
11	76	21.8	708	2 D96711	hypothetical prote
12	74	21.3	1172	2 T00665	hypothetical prote
13	73.5	21.1	291	2 S33209	extensin-like prot
14	73	21.0	299	2 T09792	proline-rich prote
15	73	21.0	794	2 I59372	cadherin 12 - huma
16	72.5	20.8	286	2 S72384	hypothetical prote
17	72.5	20.8	886	1 RGBVL3	regulatory protein
18	71.5	20.5	497	2 T27012	hypothetical prote
19	71.5	20.5	590	2 A48461	ovarian abundant m
20	70.5	20.3	225	2 T09964	extensin CYC15 pre
21	70	20.1	416	2 JU0465	extensin precursor
22	70	20.1	785	2 I50180	cadherin-7 - chick
23	69	19.8	308	2 F86147	hypothetical prote
24	69	19.8	318	2 S3358	extensin - common
25	69	19.8	1473	2 A35186	salivary agglutini
26	68.5	19.7	214	2 T10737	extensin-like cell
27	68.5	19.7	214	2 T09854	proline-rich cell
28	68.5	19.7	250	2 E81734	conserved hypothet
29	68.5	19.7	449	2 F75558	hypothetical prote

30	68.5	19.7	737	2 S28030	DNA-binding protei
31	68.5	19.7	1843	2 S18803	collagen alpha 1(V
32	68	19.5	139	2 S61885	extensin precursor
33	68	19.5	499	2 S09880	hypothetical prote
34	68	19.5	710	2 D96728	hypothetical prote
35	68	19.5	1607	2 T21982	hypothetical prote
36	67.5	19.4	510	2 S10724	surface protein pr
37	67.5	19.4	758	2 S54522	hypothetical prote
38	67.5	19.4	823	2 AD1935	general secretion
39	67.5	19.4	1712	2 A38261	masking protein pr
40	67	19.3	450	2 T17234	hypothetical prote
41	67	19.3	1257	2 S28764	neurocan precursor
42	67	19.3	3942	2 T42730	Bassoon protein -
43	66.5	19.1	372	2 T32370	hypothetical prote
44	66.5	19.1	375	2 T18178	hypothetical prote
45	66.5	19.1	1012	2 I53172	RAE-28 - mouse

## ALIGNMENTS

### RESULT 1

SAVLWD

large surface antigen - duck hepatitis virus  
N:Contains: major surface antigen; middle surface antigen

C:Species: duck hepatitis virus, DHBV

C>Date: 20-Sep-1984 #sequence\_revision 20-Sep-1984 #text\_change 13-Mar-1997

C:Accession: A03710; S12845

R:Mandart, E.; Kay, A.; Galibert, F.

J. Virol. 49, 782-792, 1984

A:Title: Nucleotide sequence of a cloned duck hepatitis B virus genome: comparison wi

A:Reference number: A92997; MUID:84138772; PMID:6699938

A:Accession: A03710

A:Molecule type: DNA

A:Residues: 1-364 <MAN>

A:Cross-references: GB:K01834

R:Mattes, F.; Tong, S.; Teubner, K.; Blum, H.E.

Nucleic Acids Res. 18, 6140, 1990

A:Title: Complete nucleotide sequence of a German duck hepatitis B virus.

A:Reference number: S12843; MUID:91045092; PMID:2235507

A:Accession: S12845

A>Status: translation not shown

A:Molecule type: DNA

A:Residues: 36-364 <NAT>

A:Cross-references: EMBL:X12798

C:Genetics:

A:Gene: pre-S1/pre-S2/S

C:Superfamily: hepatitis B virus surface antigen

C:Keywords: glycoprotein; surface antigen

F;89-364/Product: middle surface antigen (gene pre-S2/S) #status predicted <D>

F;198-364/Product: major surface antigen (gene S) #status predicted <MSA>

F;32.170.296/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 348; DB 1; Length 364;

Best Local Similarity 100.0%; Pred. No. 1.8e-28;

Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EAFRRYQERPEPTTTPPSPPQWKLPQDDPLGNQSLLETHPLXQSEPAVPVTKPP 60

Db 134 EAFRRYQERPEPTTTPPSPPQWKLPQDDPLGNQSLLETHPLXQSEPAVPVTKPP 193

QY 61 LKKK 64

Db 194 LKKK 197

### RESULT 2

SAVLWE

large surface antigen - duck hepatitis virus (strain China)

N:Contains: major surface antigen; middle surface antigen

C:Species: duck hepatitis virus, DHBV

C>Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 16-Jul-1999

C:Accession: S12842

R:Tong, S.; Mattes, F.; Teubner, K.; Blum, H.E.  
Nucleic Acids Res. 18, 6139, 1990  
A:Title: Complete nucleotide sequence of a Chinese duck hepatitis B virus.  
A:Reference number: S12840; MUID:91045091; PMID:2235506  
A:Accession: S12842  
A>Status: translation not shown  
A:Molecule type: DNA  
A:Residues: 1-365 <TONG>  
A:Cross-references: GB:M21953; NID:g325435; PIDN:AAA45746.1; PID:g325438  
C:Genetics:  
A:Gene: pre-S1/pre-S2/S  
C:Superfamily: hepatitis B virus surface antigen  
C:Keywords: glycoprotein; surface antigen  
F:89-366/Product: middle surface antigen (gene pre-S2/S) #status predicted <DSA>  
F:199-365/Product: major surface antigen (gene S) #status predicted <MSA>  
F:297/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 87.2%; Score 303.5; DB 1; Length 365;  
Best Local Similarity 89.2%; Pred. No. 6.9e-24;  
Matches 58; Conservative 3; Mismatches 3; Indels 1; Gaps 1;

1 EAFRYQERPPETTTIPSSPPQKLOPDDPLLGNSLLETHLYOS-EPAPVPIKTP 59  
|||||  
Db 134 EAFRYQERPPETTTIPSSPPQKLOPDDPLLGNSLLETHLYONPEAPVPIKTP 193  
|||||

Qy 60 PLKKK 64  
|||||

Db 194 PLKKK 198

RESULT 3  
SAVLWD  
large surface antigen - duck hepatitis virus (strain S31)  
N:Contains: major surface antigen; middle surface antigen  
C:Species: duck hepatitis virus, DHBV  
A:Note: host (Shanghai white duck)  
C:Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 04-Oct-1996  
C:Accession: D33746  
R:Uchida, M.; Esumi, M.; Shikata, T.  
Virology 173, 600-606, 1989  
A:Title: Molecular cloning and sequence analysis of duck hepatitis B virus genomes of a Chinese J. Virol. 10, 1-7, 1994  
A:Reference number: A33746; MUID:90085807; PMID:2596031  
A:Accession: D33746  
A:Molecule type: DNA  
A:Residues: 1-366 <UCH>  
A:Cross-references: GB:M32990  
C:Genetics:  
A:Gene: pre-S1/pre-S2/S  
C:Superfamily: hepatitis B virus surface antigen  
C:Keywords: glycoprotein; surface antigen  
F:89-366/Product: middle surface antigen (gene pre-S2/S) #status predicted <DSA>  
F:200-366/Product: major surface antigen (gene S) #status predicted <MSA>  
F:170,298/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 86.9%; Score 302.5; DB 1; Length 366;  
Best Local Similarity 87.7%; Pred. No. 8.8e-24;  
Matches 57; Conservative 3; Mismatches 3; Indels 1; Gaps 1;

1 EAFRYQERPPETTTIPSSPPQKLOPDDPLLGNSLLETHLYOS-EPAPVPIKTP 59  
|||||  
Db 134 EAFRYQERPPETTTIPSSPPQKLOPDDPLLGNSLLETHLYONPEAPVPIKTP 193  
|||||

Qy 60 PLKKK 64  
|||||

Db 194 PLKKK 198

RESULT 4  
SAVLBD  
large surface antigen - duck hepatitis virus (strain S5)  
N:Contains: major surface antigen; middle surface antigen  
C:Species: duck hepatitis virus, DHBV  
A:Note: host (Shanghai brown duck)

C:Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 04-Oct-1996  
C:Accession: C33746  
R:Uchida, M.; Esumi, M.; Shikata, T.  
Virology 173, 600-606, 1989  
A:Title: Molecular cloning and sequence analysis of duck hepatitis B virus genomes of a Chinese J. Virol. 10, 1-7, 1994  
A:Reference number: A33746; MUID:90085807; PMID:2596031  
A:Accession: C33746  
A:Molecule type: DNA  
A:Residues: 1-366 <UCH>  
A:Cross-references: GB:M32990  
C:Genetics:  
A:Gene: pre-S1/pre-S2/S  
C:Superfamily: hepatitis B virus surface antigen  
C:Keywords: glycoprotein; surface antigen  
F:89-366/Product: middle surface antigen (gene pre-S2/S) #status predicted <DSA>  
F:200-366/Product: major surface antigen (gene S) #status predicted <MSA>  
F:170,298/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 85.5%; Score 297.5; DB 1; Length 366;  
Best Local Similarity 87.7%; Pred. No. 2.9e-23;  
Matches 57; Conservative 3; Mismatches 4; Indels 1; Gaps 1;

1 EAFRYQERPPETTTIPSSPPQKLOPDDPLLGNSLLETHLYOS-EPAPVPIKTP 59  
|||||  
Db 134 EAFRYQERPPETTTIPTPTPKLQPGDDPLLEKSLLETHLYONPEAPVPIKTP 193  
|||||

Qy 60 PLKKK 64  
|||||

Db 194 PLKKK 198

RESULT 5  
JC1095  
Pre-S protein - duck hepatitis virus  
C:Species: duck hepatitis virus, DHBV  
C:Date: 27-Aug-1995 #sequence\_revision 27-Oct-1995 #text\_change 09-May-1997  
C:Accession: JC1095  
R:Ma, Z.M.; Li, B.L.; Xiong, S.D.; Wen, Y.M.  
Chinese J. Virol. 10, 1-7, 1994  
A:Title: High expression of duck hepatitis B virus pre-S antigen in Escherichia coli  
A:Reference number: JC1095  
A:Accession: JC1095  
A:Molecule type: DNA  
A:Residues: 1-158 <MAZ>  
C:Genetics:  
A:Gene: pre-S  
C:Superfamily: hepatitis B virus surface antigen  
C:Keywords: surface antigen

Query Match 48.6%; Score 169; DB 2; Length 158;  
Best Local Similarity 100.0%; Pred. No. 1.9e-10;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 EAFRYQERPPETTTIPSSPPQKLOPG 30  
|||||  
Db 127 EAFRYQERPPETTTIPSSPPQKLOPG 156  
|||||

RESULT 6  
SAVLHH  
large surface antigen - heron hepatitis virus  
N:Contains: major surface antigen; middle surface antigen  
C:Species: heron hepatitis virus, HHV  
A:Note: host Ardea cinerea (gray heron)  
C:Date: 30-Sep-1989 #sequence\_revision 30-Sep-1989 #text\_change 16-Jul-1999  
C:Accession: B30082  
R:Sprenkel, R.; Kaleta, E.F.; Will, H.  
J. Virol. 62, 3832-3839, 1988  
A:Title: Isolation and characterization of a hepatitis B virus endemic in herons.  
A:Reference number: A93037; MUID:88333160; PMID:83418788  
A:Accession: B30082  
A:Molecule type: DNA  
A:Residues: 1-335 <SPR>

```
Query Match      21.8%; Score 76; DB 2; Length 224;
Best Local Similarity 40.0%; Pred. No. 1.1;
```

hypothetical protein F74J5.8 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change  
C:Accession: D96711  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kae-  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.



Search completed: January 2, 2003, 13:45:36  
Job time : 10.5484 secs

This Page Blank (uspic)

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2003, 13:40:48 ; Search time 5.16129 Seconds  
(without alignments)  
514.306 Million cell updates/sec

Title: US-09-818-066-34\_COPY\_98\_161

Perfect score: 348

Sequence: 1 EAFRRYQERPPPTTIPPS.....PLYQSEPAVPVITPPLKKK 64

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	348	100.0	328	1	VMSA_HPBDU
2	303.5	87.2	365	1	VMSA_HPBDC
3	302.5	86.9	366	1	VMSA_HPBDM
4	237.5	85.5	366	1	VMSA_HPBDB
5	136.5	39.2	335	1	VMSA_HPBHE
6	80.5	23.1	328	1	ZIPA_SALTY
7	73	21.0	794	1	CADC_HUMAN
8	73	21.0	2715	1	TRX2_HUMAN
9	72.5	20.8	886	1	LEUR_YEAST
10	71.5	20.5	590	1	QAM_ASCSU
11	71	20.4	785	1	CAD7_HUMAN
12	70.5	20.3	708	1	HPS4_HUMAN
13	70.5	20.3	920	1	AD19_MOUSE
14	70	20.1	785	1	CAD7_CHICK
15	69	19.8	1500	1	SSP5_STRGN
16	68.5	19.7	895	1	DAG1_BOVIN
17	67.5	19.4	327	1	HUNB_MANSE
18	67.5	19.4	758	1	YMB8_YEAST
19	67.5	19.4	1712	1	LTBI_RAT
20	67	19.3	565	1	MOT8_MOUSE
21	67	19.3	1257	1	PCGN_RAT
22	66	19.0	751	1	FPL_MYTGA
23	65.5	18.8	299	1	PAF_HUMAN
24	65.5	18.8	346	1	PRF1_LYCES
25	65.5	18.8	728	1	P85B_HUMAN
26	65.5	18.8	778	1	MGD1_HUMAN
27	65.5	18.8	799	1	CAD8_HUMAN
28	65.5	18.8	799	1	CAD8_MOUSE
29	65.5	18.8	799	1	CAD8_RAT
30	65.5	18.8	893	1	DAG1_MOUSE
31	65.5	18.8	895	1	DAG1_HUMAN
32	65.5	18.8	895	1	DAG1_RABIT
33	65.5	18.8	1189	1	HAIR_HUMAN

RESULT 1

ID	VMSA_HPBDU	STANDARD;	PRT;	328 AA.
AC	P03145;			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	01-JUL-1993 (Rel. 26, Last sequence update)			
DT	01-FEB-1994 (Rel. 28, Last annotation update)			
DE	Major surface antigen precursor.			
GN	S.			
OS	Duck hepatitis B virus (DHBV).			
OC	Viruses; Retroviral viruses; Hepadnaviridae; Avihepadnavirus.			
OX	NCBI_TaxID=12639;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=84138772; PubMed=6699938;			
RA	Mandart E., Kay A., Galibert F.;			
RT	"Nucleotide sequence of a cloned duck hepatitis B virus genome: comparison with woodchuck and human hepatitis B virus sequences.";			
RL	J. Virol. 49:782-792(1984).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Isolate DHBV FL-6;			
RX	MEDLINE=91045092; PubMed=2235507;			
RA	Mattes F., Tong S., Teubner K., Blum H.E.;			
RT	"Complete nucleotide sequence of a German duck hepatitis B virus.";			
RL	Nucleic Acids Res. 18:6140-6140(1990).			
RN	[3]			
RP	MYRISTOYLATION.			
RX	MEDLINE=91135002; PubMed=1994583;			
RA	Macrae D.R., Bruss V., Ganem D.;			
RT	"Myristylation of a duck hepatitis B virus envelope protein is essential for infectivity but not for virus assembly.";			
RL	Virology 181:359-363(1991).			
CC	-1- PPM: MYRISTOYLATION CONTRIBUTES IMPORTANTLY TO DHBV INFECTIVITY. IT IS MOST LIKELY REQUIRED FOR AN EARLY STEP OF THE LIFE CYCLE INVOLVING THE ENTRY OR UNCOATING OF VIRUS PARTICLES.			
CC	-----			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; X12798; CAB57224.1;			
DR	PIR; A03710; SAVID.			
DR	PIR; S12845; S12845.			
DR	InterPro: IPR000349; Hepvir_surfacg.			
DR	Pfam: PF00695; vmsa; 1.			
KW	Antigen; Myristate; Envelope protein; Lipoprotein.			
PROPEP	1 161			
FT	CHAIN 162 328			
FT	LIPID 2 2			
FT	CARBOHYD. 134 134			
FT	CARBOHYD. 260 260			

```

GN      Duck hepatitis B virus (White Shanghai duck isolate S31) (DHBV).
OC      Viruses; Retroid viruses; Hepadnaviridae; Avihepadnavirus.
OX      NCBI_TaxID=10440;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=90085807; PubMed=2596031;
RA      Uchida M., Esumi M., Shikata T.;
RT      "Molecular cloning and sequence analysis of duck hepatitis B virus
RT      genomes of a new variant isolated from Shanghai ducks.";
RL      Virology 173:600-606(1989).
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC      or send an email to license@isb-sib.ch).
-----
DR      EMBL; M32991; AAA45752.1; ALT_INIT.
DR      PIR; D33746; SAVLMD.
DR      InterPro; IPR000349; Hepvir_surfAg.
DR      Pfam; PF00695; VMSA; 1.
KW      Antigen.
FT      PROPEP      1      199
FT      CHAIN       200      366      MAJOR SURFACE ANTIGEN.
FT      CARBOHYD    170      170      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD    298      298      N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ      SEQUENCE    366 AA; 40858 MW; CF60E78B72FCD52 CRC64;
-----
Query Match      86.9%; Score 302.5; DB 1; Length 366;
Best Local Similarity 87.7%; Pred. No. 5.2e-23;
Matches 57; Conservative 4; Mismatches 3; Indels 1; Gaps 1;
-----
QY      1 EAFRYQEERPPETTTTPSSPPQWKLPQGDPLLGNSLLETHPLYQS-EPAPVPIKTP 59
Db      134 EAFRYQEERPPETTTTPPTPTPKLQPGDPLLENKSLLETHPLYQNPEAPVPIKTP 193
QY      60 PLKKK 64
Db      194 PLRKK 198
-----
RESULT 4
VMSA_HPBDB
ID      VMSA_HPBDB      STANDARD;      PRT;      366 AA.
AC      PL7194;
DT      01-AUG-1990 (Rel. 15, Created)
DT      01-MAY-1991 (Rel. 18, Last sequence update)
DT      01-MAY-1991 (Rel. 18, Last annotation update)
DE      Major surface antigen precursor.
GN      S.
OC      Duck hepatitis B virus (brown Shanghai duck isolate S5) (DHBV).
OX      Viruses; Retroid viruses; Hepadnaviridae; Avihepadnavirus.
OX      NCBI_TaxID=10439;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=90085807; PubMed=2596031;
RA      Uchida M., Esumi M., Shikata T.;
RT      "Molecular cloning and sequence analysis of duck hepatitis B virus
RT      genomes of a new variant isolated from Shanghai ducks.";
RL      Virology 173:600-606(1989).
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC      or send an email to license@isb-sib.ch).
-----
DR      EMBL; M32990; AAA45755.1; ALT_INIT

```



```
DR PIR: C33746; SAVLBD.  
DR InterPro: IPR000349; Hepvir_surfac.  
DR Pfam: PF00695; VMSA; 1.  
KW Antigen.  
FT PROPEP 1 199  
FT CHAIN 200 366 MAJOR SURFACE ANTIGEN.  
FT CARBOHYD 170 170 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 298 298 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 366 AA; 40897 MW; 5872879A182EFF38 CRC64;  
  
Query Match 85.5%; Score 297.5; DB 1; Length 366;  
Best Local Similarity 87.7%; Pred. No. 1.6e-22;  
Matches 57; Conservative 3; Mismatches 4; Indels 1; Gaps 1;  
  
QY 1 EAFRRYQERPPETTTTPSPQWKLQGDPLLNQSLLETHPLYS-EPAPVPVTKP 59  
DB 134 EAFRRYQERPPETTTTPSPQWKLQGDPLLNQSLLETHPLYS-EPAPVPVTKP 193  
  
QY 60 PLKKK 64  
DB 194 PLKKK 198  
  
RESULT 5  
VMSA_HPBHE STANDARD; PRT; 335 AA.  
AC P13847; Q67853;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Major surface antigen precursor.  
GN S.  
OS Heron hepatitis b virus.  
OC Viruses; Retroid viruses; Hepadnaviridae; Avihepadnavirus.  
OX NCBI_TaxID=28300;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88333160; PubMed=3418788;  
RA Sprengel R., Kaleta E.F., Will H.;  
RT Isolation and characterization of a hepatitis B virus endemic in  
herons.*;  
RL J. Virol. 62:3832-3839(1988).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC  
CC EMBL; M22056; AAA45739.1; -;  
DR EMBL; M22056; AAA45740.1; ALT_INIT.  
DR PIR; B30082; SAVLHH.  
DR InterPro: IPR000349; Hepvir_surfac.  
DR Pfam: PF00695; VMSA; 1.  
KW Antigen.  
FT PROPEP 1 166  
FT CHAIN 167 335 MAJOR SURFACE ANTIGEN.  
FT CARBOHYD 265 265 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 335 AA; 37218 MW; E90F7DA6A0623354 CRC64;  
  
Query Match 39.2%; Score 136.5; DB 1; Length 335;  
Best Local Similarity 40.8%; Pred. No. 1.2e-06;  
Matches 29; Conservative 10; Mismatches 19; Indels 13; Gaps 3;  
  
QY 1 EAFRRYQERPPETTTTPP-----SSPQWKLQGDPLLNQSLLETHPLYSQSPAV 53  
DB 100 EFPQYQENRKPAPETAPPPITELHAEPPQWKRISP-EDPLKAKALIPV-----KEPEV 153  
  
QY 54 PVIKTPPLKKK 64  
DB 154 PILKVPKLTNK 164
```

```
RESULT 6  
ZIPA_SALTY STANDARD; PRT; 328 AA.  
ID ZIPA_SALTY  
AC P55894;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Cell division protein zipa.  
DE Cell division protein zipa.  
GN ZIPA OR STM2428 OR STV2664.  
OS Salmonella typhimurium, and  
OS Salmonella typhi.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Salmonella.  
OX NCBI_TaxID=602, 601;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;  
RX MEDLINE=21534948; PubMed=11677609;  
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
RA Waterston R., Wilson R.K.;  
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium  
LT2.*";  
RL Nature 413:852-856(2001).  
RN [2]  
RP SEQUENCE OF 1-32 FROM N.A.  
RC SPECIES=S.typhimurium; STRAIN=LT2;  
RX MEDLINE=88257033; PubMed=3290198;  
RA Byrne C.R., Monroe R.S., Ward K.A., Kredich N.M.;  
RT "DNA sequences of the cysK regions of Salmonella typhimurium and  
Escherichia coli and linkage of the cysK regions to ptsH.*";  
RL J. Bacteriol. 170:3150-3157(1988).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC SPECIES=S.typhi; STRAIN=CT18;  
RX MEDLINE=21534947; PubMed=11677608;  
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,  
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,  
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,  
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,  
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,  
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,  
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,  
RA Whitehead S., Barrall B.G.;  
RT "Complete genome sequence of a multiple drug resistant Salmonella  
enterica serovar Typhi CT18.*";  
RL Nature 413:848-852(2001).  
CC -1- FUNCTION: Interacts directly with the cell division protein ftsZ.  
CC Probable receptor for the septal ring structure, may anchor it  
CC to the inner-membrane (By similarity).  
CC -1- SUBCELLULAR LOCATION: Type IB membrane protein. Inner membrane (By  
CC similarity).  
CC -1- SIMILARITY: BELONGS TO THE ZIPA FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC  
CC EMBL; AE008809; AAL21322.1; -;  
DR EMBL; M21450; -; NOT_ANNOTATED_CDS.  
DR EMBL; AL627274; CAD07660.1; -;  
DR StyGene; SG10748; zipa.  
KW Cell division; Septation; Transmembrane; Inner membrane;  
KW Complete proteome.  
FT DOMAIN 1 6 PERIPLASMIC (POTENTIAL).
```

**PROSITE**; PS50268; CADHERIN\_2; 5.  
Signal. signal  
**SIGNAL**:  
**PROPEP**:  
**CHAIN**:  
**DOMAIN**:  
**TRANSMEM**:  
**DOMAIN**:  
**DOMAIN**:  
**DOMAIN**:  
**DOMAIN**:  
**DOMAIN**:  
**DOMAIN**:  
**DOMAIN**:  
**CADHERIN**:  
**CADHERIN**:  
**CADHERIN**:  
**CADHERIN**:  
**CADHERIN**:  
**N-LINKED**:  
**N-LINKED**:  
**N-LINKED**:  
**N-LINKED**:  
**V->**:  
**E->**:  
**G->**:  
**H->**:  
**I->**:  
**T->**:  
**A->**:

POTENTIAL.  
POTENTIAL.  
BRAIN-CADHERIN.  
EXTRACELLULAR (POTENTIAL).  
POTENTIAL.  
CYTOPLASMIC (POTENTIAL).  
CADHERIN 1.  
CADHERIN 2.  
CADHERIN 3.  
CADHERIN 4.  
CADHERIN 5.  
N-LINKED (GLCNAC. .) (POTENTIAL).  
N-LINKED (GLCNAC. .) (POTENTIAL).  
N-LINKED (GLCNAC. .) (POTENTIAL).  
N-LINKED (GLCNAC. .) (POTENTIAL).  
V -> M (IN REF. 2).  
E -> D (IN REF. 2).  
G -> S (IN REF. 2).  
H -> D (IN REF. 2).  
I -> Y (IN REF. 2).  
T -> T (IN REF. 2).

Query Match      21.0%; Score 73; DB 1; Length 794;  
Best Local Similarity    33.3%; Pred. No. 5,7;  
Matches    19; Conservative    6; Mismatches    22; Indels    10; Gaps    2;

**OY**    12 PPTTTTPSSPP-----QWKLGQDDPLLGNQ-----SLLETHLPXQSEPAVPVKIT 58  
| : | : || | | | | | | | | | : | : | | : | : | : |  
**Ddb**    164 PVVATVPENSPGVAYLVQKVAADDPTGYNSARVVSIILQGOPYFSIDPKTGIVRT 220

**RESULT 8**

**TXRX2\_HUMAN**  
TXRX2\_HUMAN STANDARD; PRT; 2715 AA.  
Q9UNM6; Q9UKZ5; Q9S836; Q9Y669; Q9Y668; O15022; Q96GP2; Q96IP3;  
16-OCT-2001 (Rel. 40, Created)  
16-OCT-2001 (Rel. 40, Last sequence update)  
15-JUN-2002 (Rel. 41, Last annotation update)  
Trithorax homolog 2 (Mixed lineage leukemia gene homolog 2 protein).  
TXRX2 OR HRX2 OR MLL2 OR MLL4 OR KTAO304.  
Homo sapiens (Human).  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
NCBI\_Taxid=9606;  
[1]  
SEQUENCE FROM N.A. (LONG ISOFORM).  
Angrand P.O., Valvatne H., Jeanmougin F., Adamson A., van der Hoeven F., Olsen L., Tekotte H., Huang N., Poeh O., Lamerdin J., Chambon P., Losson R., Stewart A., Asaland R.; "Mamalian trithorax- and ASHL-like proteins: putative chromatin regulators which contain PHD fingers and SET domains." Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.

[2]  
SEQUENCE FROM N.A. (LONG ISOFORM).  
Lamerdin J.E., McCreedy P.M., Adamson A.W., Burkhardt-Schultz K., Garcia E., Kyle A., Ramirez M., Stillwagen S., Garnes J., Panganan L., Bruce R., Quan G., Montgomery M., Ow D., Kobayashi A., Olsen A.O., Carrano A.V.; "Sequence analysis of a 1 Mb region in human 19ql3.1."; Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.

[3]  
SEQUENCE OF l1l-2715 FROM N.A. (LONG ISOFORM).  
Tissue-testis, and Leukocyte:  
MEDLINE=20105772; PubMed=10637508;  
Huntman D.G., Chin S.-F., Muller M., Batley S.J., Collins V.P., Wiedemann L.M., Aparicio S., Caldas C.; "MLL2, the second human homolog of the Drosophila trithorax gene, maps to 19ql3.1 and is amplified in solid tumor cell lines." Oncogene 18:7975-7984(1999).

[4] SEQUENCE OF 816-2715 FROM N.A. (LONG ISOFORM).  
 TSU-Brain;  
 MEDLINE-97349984; PubMed-9205841;  
 Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,  
 Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;  
 "Prediction of the coding sequences of unidentified human genes. VII.  
 The complete sequences of 100 new cDNA clones from brain which can  
 code for large proteins in vitro.";  
 DNA Res. 4:141-150(1997).  
 [5]  
 SEQUENCE OF 1918-2715 FROM N.A.  
 TSU-Brain, and Skin;  
 Strausberg R.;  
 Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 [6]  
 PARTIAL SEQUENCE FROM N.A. (LONG AND TRUNCATED ISOFORMS).  
 TSU-placenta, and Bone marrow;  
 MEDLINE-99339983; PubMed-10409430;  
 Fitzgerald K.T., Diaz M.O.;  
 "MLL2: A new mammalian member of the trx/MLL family of genes.";  
 Genomics 59:187-192(1999).  
 [7]  
 [8]  
 [9]  
 [10]  
 [11]  
 [12]  
 [13]  
 [14]  
 [15]  
 [16]  
 [17]  
 [18]  
 [19]  
 [20]  
 [21]  
 [22]  
 [23]  
 [24]  
 [25]  
 [26]  
 [27]  
 [28]  
 [29]  
 [30]  
 [31]  
 [32]  
 [33]  
 [34]  
 [35]  
 [36]  
 [37]  
 [38]  
 [39]  
 [40]  
 [41]  
 [42]  
 [43]  
 [44]  
 [45]  
 [46]  
 [47]  
 [48]  
 [49]  
 [50]  
 [51]  
 [52]  
 [53]  
 [54]  
 [55]  
 [56]  
 [57]  
 [58]  
 [59]  
 [60]  
 [61]  
 [62]  
 [63]  
 [64]  
 [65]  
 [66]  
 [67]  
 [68]  
 [69]  
 [70]  
 [71]  
 [72]  
 [73]  
 [74]  
 [75]  
 [76]  
 [77]  
 [78]  
 [79]  
 [80]  
 [81]  
 [82]  
 [83]  
 [84]  
 [85]  
 [86]  
 [87]  
 [88]  
 [89]  
 [90]  
 [91]  
 [92]  
 [93]  
 [94]  
 [95]  
 [96]  
 [97]  
 [98]  
 [99]  
 [100]  
 [101]  
 [102]  
 [103]  
 [104]  
 [105]  
 [106]  
 [107]  
 [108]  
 [109]  
 [110]  
 [111]  
 [112]  
 [113]  
 [114]  
 [115]  
 [116]  
 [117]  
 [118]  
 [119]  
 [120]  
 [121]  
 [122]  
 [123]  
 [124]  
 [125]  
 [126]  
 [127]  
 [128]  
 [129]  
 [130]  
 [131]  
 [132]  
 [133]  
 [134]  
 [135]  
 [136]  
 [137]  
 [138]  
 [139]  
 [140]  
 [141]  
 [142]  
 [143]  
 [144]  
 [145]  
 [146]  
 [147]  
 [148]  
 [149]  
 [150]  
 [151]  
 [152]  
 [153]  
 [154]  
 [155]  
 [156]  
 [157]  
 [158]  
 [159]  
 [160]  
 [161]  
 [162]  
 [163]  
 [164]  
 [165]  
 [166]  
 [167]  
 [168]  
 [169]  
 [170]  
 [171]  
 [172]  
 [173]  
 [174]  
 [175]  
 [176]  
 [177]  
 [178]  
 [179]  
 [180]  
 [181]  
 [182]  
 [183]  
 [184]  
 [185]  
 [186]  
 [187]  
 [188]  
 [189]  
 [190]  
 [191]  
 [192]  
 [193]  
 [194]  
 [195]  
 [196]  
 [197]  
 [198]  
 [199]  
 [200]  
 [201]  
 [202]  
 [203]  
 [204]  
 [205]  
 [206]  
 [207]  
 [208]  
 [209]  
 [210]  
 [211]  
 [212]  
 [213]  
 [214]  
 [215]  
 [216]  
 [217]  
 [218]  
 [219]  
 [220]  
 [221]  
 [222]  
 [223]  
 [224]  
 [225]  
 [226]  
 [227]  
 [228]  
 [229]  
 [230]  
 [231]  
 [232]  
 [233]  
 [234]  
 [235]  
 [236]  
 [237]  
 [238]  
 [239]  
 [240]  
 [241]  
 [242]  
 [243]  
 [244]  
 [245]  
 [246]  
 [247]  
 [248]  
 [249]  
 [250]  
 [251]  
 [252]  
 [253]  
 [254]  
 [255]  
 [256]  
 [257]  
 [258]  
 [259]  
 [260]  
 [261]  
 [262]  
 [263]  
 [264]  
 [265]  
 [266]  
 [267]  
 [268]  
 [269]  
 [270]  
 [271]  
 [272]  
 [273]  
 [274]  
 [275]  
 [276]  
 [277]  
 [278]  
 [279]  
 [280]  
 [281]  
 [282]  
 [283]  
 [284]  
 [285]  
 [286]  
 [287]  
 [288]  
 [289]  
 [290]  
 [291]  
 [292]  
 [293]  
 [294]  
 [295]  
 [296]  
 [297]  
 [298]  
 [299]  
 [300]  
 [301]  
 [302]  
 [303]  
 [304]  
 [305]  
 [306]  
 [307]  
 [308]  
 [309]  
 [310]  
 [311]  
 [312]  
 [313]  
 [314]  
 [315]  
 [316]  
 [317]  
 [318]  
 [319]  
 [320]  
 [321]  
 [322]  
 [323]  
 [324]  
 [325]  
 [326]  
 [327]  
 [328]  
 [329]  
 [330]  
 [331]  
 [332]  
 [333]  
 [334]  
 [335]  
 [336]  
 [337]  
 [338]  
 [339]  
 [340]  
 [341]  
 [342]  
 [343]  
 [344]  
 [345]  
 [346]  
 [347]  
 [348]  
 [349]  
 [350]  
 [351]  
 [352]  
 [353]  
 [354]  
 [355]  
 [356]  
 [357]  
 [358]  
 [359]  
 [360]  
 [361]  
 [362]  
 [363]  
 [364]  
 [365]  
 [366]  
 [367]  
 [368]  
 [369]  
 [370]  
 [371]  
 [372]  
 [373]  
 [374]  
 [375]  
 [376]  
 [377]  
 [378]  
 [379]  
 [380]  
 [381]  
 [382]  
 [383]  
 [384]  
 [385]  
 [386]  
 [387]  
 [388]  
 [389]  
 [390]  
 [391]  
 [392]  
 [393]  
 [394]  
 [395]  
 [396]  
 [397]  
 [398]  
 [399]  
 [400]  
 [401]  
 [402]  
 [403]  
 [404]  
 [405]  
 [406]  
 [407]  
 [408]  
 [409]  
 [410]  
 [411]  
 [412]  
 [413]  
 [414]  
 [415]  
 [416]  
 [417]  
 [418]  
 [419]  
 [420]  
 [421]  
 [422]  
 [423]  
 [424]  
 [425]  
 [426]  
 [427]

RP SEQUENCE FROM N.A.  
 RC STRAIN-S288C / AB972;  
 RA Johnson M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,  
 RA Favellio A., Fulton L., Gattung S., Greco T., Kirsten J.,  
 RA Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,  
 RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,  
 RA Mardis E., Menezes S., Miller N., Nhan M., Pauley A., Peluso D.,  
 RA Rifkin L., Riles L., Raich A., Trevaskis E., Vignati D.,  
 RA Wilcox L., Wohlman P., Vaudin M., Wilson R., Waterston R.,  
 RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: FACTOR FOR CONTROL OF RNA LEVELS OF A GROUP OF  
 CC LEUCINE-SPECIFIC GENES.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- SIMILARITY: CONTAINS 1 ZN(2)-CYS(6), FUNGAL-TYPE BINUCLEAR  
 CC CLUSTER DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; Y00360; CAA68438.1; -  
 DR EMBL; M17222; AAB34741.1; -  
 DR EMBL; U22382; AAB67526.1; -  
 DR PIR; S00638; RGYL3.  
 DR HSSP; P07272; 1PVI.  
 DR TRANSFAC; T00465; -  
 DR SGD; S0004443; LEU3.  
 DR InterPro; IPR001138; Fungi\_Tn.  
 DR Pfam; PF00172; Zn\_c1us; 1.  
 DR SMART; SM00066; GAL4; 1.  
 DR PROSITE; PS00463; ZN2\_Cy6\_FUNGAL\_1; 1.  
 DR PROSITE; PS00463; ZN2\_Cy6\_FUNGAL\_2; 1.  
 KW Transcription regulation; Activator; DNA-binding; Nuclear protein;  
 KW Zinc; Metal-binding; Leucine biosynthesis.  
 FT DNA\_BIND 37 67 ZN(2)-CYS(6), FUNGAL-TYPE.  
 FT DOMAIN 67 697 ASP/GLU-RICH (ACIDIC).  
 FT CONFLICT 504 504 M -> I (IN REF. 2).  
 FT SEQUENCE 886 AA; 100153 MW; C5F8EDCD326B1EBB CRC64;  
 SQ  
 Query Match 20.8%; Score 72.5; DB 1; Length 886;  
 Best Local Similarity 39.1%; Pred. No. 7.2;  
 Matches 18; Conservative 5; Mismatches 20; Indels 3; Gaps 1;  
 OY 18 PPSPQWKLQPGDDPL---LGNOSLLETHPLYSQSEPAVPVKTTP 60  
 809 PANQPLSITQWQENTLPATQANSLLLETPIVOSNPVTTIKESP 854  
 RESULT 10  
 OAM\_ASCSU  
 ID OAM\_ASCSU STANDARD; PRT; 590 AA.  
 AC Q01456;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last annotation update)  
 DE Ovarian abundant message protein (OAM protein).  
 GN OAM.  
 OS Ascaris suum (Pig roundworm) (Ascaris lumbricoidea).  
 CC Eukaryota; Metazoa; Nematoda; Chromadorea; Ascarididae; Ascaridoidea;  
 CC Ascarididae; Ascaris.  
 CC NCBI\_TaxID=6253;  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 CC TISSUE=Ovary;  
 CC MEDLINE=93116800; PubMed=1474996;  
 CC Guindi M., Catter J., Wilson B., Gharib S., Bennett K.L.;  
 RA "An extremely abundant ovarian mRNA from the parasitic nematode  
 RT Ascaris lumbricoidea var. suum has multiple repeat motifs";  
 RT Mol. Biochem. Parasitol. 56:177-180(1992).  
 CC

CC -1- TISSUE SPECIFICITY: SOMATIC OVARIAN TISSUE.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; M94771; AAA73355.1; -  
 DR PIR; A48461; A48461.  
 DR PIR; S27773; S27773.  
 KW Repeat.  
 FT DOMAIN 66 185 20 X 6 AA TANDEM REPEATS, MOTIF 1.  
 FT DOMAIN 300 347 7 X 6 AA TANDEM REPEATS, MOTIF 2.  
 FT DOMAIN 348 413 11 X 6 AA TANDEM REPEATS, MOTIF 3.  
 FT DOMAIN 419 448 5 X 6 AA TANDEM REPEATS, MOTIF 4.  
 SQ SEQUENCE 590 AA; 62963 MW; 5BDB9D06918BEEF3 CRC64;  
 Query Match 20.5%; Score 71.5; DB 1; Length 590;  
 Best Local Similarity 33.8%; Pred. No. 5.8;  
 Matches 22; Conservative 10; Mismatches 22; Indels 11; Gaps 4;  
 QY 8 EERPP-ETTTTPSSP-----PQWKLQ---GDDPLGNOSLLETHPLYSQSPAV---PVI 56  
 Db 75 EERPPVEEQPLVEEQPLVEEQPLVEEQPLVEEQPLVEEQPLVEEQPLVEEQPLVEEQPLV 134  
 QY 57 KTPPL 61  
 Db 135 EEQPL 139  
 RESULT 11  
 CAD7\_HUMAN  
 ID CAD7\_HUMAN STANDARD; PRT; 785 AA.  
 AC Q9ULB5; Q9H157;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Cadherin-7 precursor.  
 GN CDH7 OR CDH7L1.  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 CC NCBI\_TaxID=9606;  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 CC MEDLINE=20453188; PubMed=10995570;  
 CC Kools P., Van Imschoot G., van Roy F.;  
 RT "Characterization of three novel human Cadherin genes (CDH7, CDH19,  
 RT and CDH20) clustered on chromosome 18q22-q23 and with high homology to  
 RT chicken Cadherin-7";  
 RL Genomics 68:283-295(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Shimoyama Y.;  
 RT "Identified of human classic cadherins";  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases  
 CC -1- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.  
 CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC  
 CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE  
 CC SORTING OF HETEROGENEOUS CELL TYPES.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- SIMILARITY: CONTAINS 5 CADHERIN DOMAINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC

```

CC  EMBL: AJ007611; CAC13127.1; -.
DR  EMBL: AB035301; BAA87415.1; -.
DR  HSP: P15116; INCH.
DR  Genew; HGNC:1766; CDH7.
DR  MIM: 605806; -.
DR  InterPro; IPR002126; Cadherin.
DR  InterPro; IPR000233; Cadherin_C_term.
DR  Pfam; PF00028; cadherin; 5.
DR  Pfam; PF01049; Cadherin_C_term; 2.
DR  PRINTS; PR00205; CADHERIN.
DR  SMART; SM00112; CA; 5.
DR  PROSITE; PS00232; CADHERIN_1; 3.
DR  PROSITE; PS00268; CADHERIN_2; 5.
KW  Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
KW  Signal.
FT  SIGNAL          1   27   POTENTIAL.
FT  PROPEP          28   47   POTENTIAL.
FT  CHAIN           48   785
FT  DOMAIN          28   607   CADHERIN-7.
FT  DOMAIN          608   628   EXTRACELLULAR (POTENTIAL).
FT  TRANSMEM      629   785   POTENTIAL.
FT  DOMAIN          629   785   CYTOPLASMIC (POTENTIAL).
FT  DOMAIN          49   153   CADHERIN 1.
FT  DOMAIN          154   262   CADHERIN 2.
FT  DOMAIN          263   377   CADHERIN 3.
FT  DOMAIN          378   482   CADHERIN 4.
FT  DOMAIN          482   599   CADHERIN 5.
FT  CARBOHYD        449   449   N-LINKED (GLCNAC. . .) (POTENTIAL).
FT  CARBOHYD        530   530   N-LINKED (GLCNAC. . .) (POTENTIAL).
FT  CONFLICT        23   23   G -> E (IN REF. 1).
FT  CONFLICT        32   32   S -> P (IN REF. 1).
FT  CONFLICT        591   600   AQTCAAEAYV -> TQTAMQRLC (IN REF. 1).
SQ  SEQUENCE       785 AA; 87058 MW; 0D889FFFE105279B CRC64;

Query Match          20.4%; Score 71; DB 1; Length 785;
Best Local Similarity 35.1%; Pred. No. 8.8;
Matches 20; Conservative 4; Mismatches 23; Indels 10; Gaps 2;

OY  12 PTTTTPSP-----PQMKLPQGDPLGQ-----SLLETHPLQSEPAVPVVIKT 58
DB  157 PYTAGVPEMSPVGTSGVQVATDADDDPTGNSARVVYSILQGPVFSVEPTGVIKT 213

RESULT 12
ID  HPS4_HUMAN          STANDARD;          PRT;          708 AA.
AC  Q9NQG7; Q9BLY3; Q96LK6; Q9UH38; Q9UH37;
DT  15-JUN-2002 (Rel. 41, Created)
DI  15-JUN-2002 (Rel. 41, Last sequence update)
DE  Hermansky-Pudlak syndrome 4 protein (Light-ear protein homolog).
GN  HPS4 OR KIAA1667.
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX  NCBI_Taxid=9606;
RN  [1]
RP  SEQUENCE FROM N.A. (ISOFORM 1).
RX  PubMed=11836498;
RA  Suzuki T., Li W., Zhang Q., Karim A., Novak E.K., Sviderskaya E.V.,
RA  Hill S.P., Bennett D.C., Levin A.V., Nieuwenhuis H.K., Fong C.-T.,
RA  Castellani C., Metersky B., Swank R.T., Spritz R.A.;
RT  "Hermansky-Pudlak syndrome is caused by mutations in HPS4, the human
RT  homolog of the mouse light-ear gene.";
RL  Nat. Genet. 30:321-324(2002).
RN  [2]
RP  SEQUENCE OF 204-708 FROM N.A. (ISOFORM 1).
RC  Tissue=Brain;
RX  MEDLINE=21156230; PubMed=11258795;
RA  Hirosewa M., Nagase T., Murahashi Y., Kikuno R., Ohara O.;
RT  "Identification of novel transcribed sequences on human chromosome 22
RT  by expressed sequence tag mapping.";
RL  DNA Res. 8:1-9(2001).

```

```

RN  [3]
RP  SEQUENCE FROM N.A. (ISOFORM 1).
RA  Collins J.E., Huckle E.J.;
RL  Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RN  [4]
RP  SEQUENCE FROM N.A. (ISOFORM 2).
RC  Tissue=Trachea;
RA  Suzuki O., Sasaki N., Aotsuka S., Shoji T., Ichihara T., Shiohata N.,
RA  Matsumoto K., Hirano M., Sano S., Nomura R., Yoshikawa Y.,
RA  Matsumura Y., Moriya S., Chiba E., Momiyama H., Onogawa S.,
RA  Kaeriyama S., Satoh N., Matsunawa H., Takahashi E., Katsoka R.,
RA  Kuga N., Kuroda A., Satoh I., Kamata K., Takami S., Terashima Y.,
RA  Watanabe M., Sugiyama T., Irie R., Otsuki T., Sato H., Ota T.,
RA  Wakamatsu A., Ishii S., Yamamoto J., Isono Y., Kawai-Hio Y., Saito K.,
RA  Nishikawa T., Kimura K., Yamashita H., Matsuo K., Nakamura Y.,
RA  Sekine M., Kikuchi H., Kanda K., Wagatsuma M., Murakawa K.,
RA  Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA  Suzuki Y., Sugano S., Nagahara K., Masuho Y., Nagai K., Isogai T.;
RL  "NEDO human cDNA sequencing project.";
RN  [5]
RP  Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RN  [6]
RP  SEQUENCE FROM N.A. (ISOFORM 1).
RC  Tissue=Testis;
RA  Ottenwaelder B., Obermaier B., Mewes H.-W., Gassenhuber J.,
RA  Wienmann S.;
RL  Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RN  [7]
RP  SEQUENCE FROM N.A.
RC  MEDLINE=20057165; PubMed=10591208;
RA  Dunham I., Hunt A.R., Collins J.E., Bruskewich R., Beare D.M.,
RA  Clump M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K.,
RA  Bagguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,
RA  Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,
RA  Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,
RA  Clegg S.M., Corbey N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
RA  Conroy D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,
RA  Evans G.B., Fey J.M., Fleming K., French L., Garner A.A.,
RA  Gilbert J.G.R., Goward M.E., Graham D.V., Griffiths M.N.D., Hall C.,
RA  Hall R.E., Hall-Tamlyn G., Heathcott R.W., Ho S., Holmes S.,
RA  Hunt S.E., Jones M.C., Kersey J., Kimberley A.M., King A.,
RA  Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,
RA  Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T.,
RA  Mcclay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,
RA  Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
RA  Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,
RA  Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L., Swann R.M.,
RA  Soderlund C., Spragon L., Steward C.A., Sulston J.E., Wynn R.M.,
RA  Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,
RA  Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,
RA  Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
RA  Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,
RA  Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,
RA  Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
RA  Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,
RA  Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,
RA  Phan Q., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,
RA  Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,
RA  Zhan M., Zhang G., Chissoe S., Murray J., Miller N., Minx P.,
RA  Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,
RA  Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
RA  Hinds K., Kemp K., Latreille P., Layman D., Ozerky P., Rohlfing T.,
RA  Scheet P., Walker C., Wamsley A., Wohlmann P., Pepin K., Nelson J.,
RA  Korf I., Bedell J.A., Hillier L., Mardis E., Waterston R., Wilson R.,
RA  Emanuel B.S., Shaikh T., Kurahashi H., Saitta S., Budarf M.L.,
RA  McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E., Edelmann L.,
RA  Kim U.J., Shizuya H., Simon M.I., Dumanski J.P., Peyraud M., Kedra D.,
RA  Seroussi E., Franssen I., Tapia I., Bruder C.E., O'Brien K.P.,
RA  Wilkison P., Bodenteich A., Hartman K., Hu X., Khan A.S., Lane L.,
RA  Tikhunov V., Wright H.;
RT  "The DNA sequence of human chromosome 22.";
RL  Nature 402:489-495(1999).
CC  -!- FUNCTION: May function in the pathway of organelle biogenesis.

```

CC -|- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; may be  
 CC produced by alternative splicing.  
 CC -|- DISEASE: Defects in HPS4 are the cause of Hermansky-Pudlak  
 CC syndrome type 4 (HPS4), an autosomal recessive disorder,  
 CC characterized by oculocutaneous albinism, bleeding due to platelet  
 CC storage pool deficiency, and lysosomal storage defects. This  
 CC syndrome results from defects of diverse cytoplasmic organelles  
 CC including melanosomes, platelet dense granules and lysosomes.  
 CC Ceroid storage in the lungs is associated with pulmonary fibrosis,  
 CC a common cause of premature death in individuals with HPS.  
 CC -|- DATABASE: NAME=Albinism database (ADB);  
 CC NOTE=HPS mutations;  
 CC WWW="http://www.cbc.umn.edu/tad/".  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AY043416; AAK95330.1; -  
 CC EMBL; AB051454; BAB33337.1; -  
 CC EMBL; AL365512; CAB97208.1; -  
 CC EMBL; AK057648; BAB15340.1; -  
 CC EMBL; AL713795; CAD28549.1; -  
 CC EMBL; Z99714; CAB62927.1; ALT\_INIT.  
 CC EMBL; Z99714; CAB62928.1; ALT\_TERM.  
 CC Genbank; HGNC:15844; HPS4.  
 CC MIM; 606682; -  
 CC MIM; 203300; -  
 CC Hermansky-Pudlak syndrome; Albinism; Disease mutation;  
 KW Alternative splicing.  
 FT VARSPLIC 236 236 G -> GKWLMSFKNRVTHQNPNG (IN ISOFORM 2).  
 FT VARSPLIC 463 510 RTREPLLRLDPCQGNKLPTEQGLDEVDGVCESHA  
 FT VARSPLIC 463 510 FGLECS -> KSVIFSGCGKSDTQLCVPLGVWLEPQ  
 FT VARSPLIC 511 708 FYIVIVRHFFDSMQTAG (IN ISOFORM 2).  
 FT CONFLICT 3 14 MISSING (IN ISOFORM 2).  
 FT CONFLICT 204 237 TSTPEAKSASW -> PICSLAR (IN REF. 3).  
 FT CONFLICT 229 229 QENNVVPSANFCVCLSETPYRGCPACTWS (IN REF.  
 FT CONFLICT 430 430 E -> G (IN REF. 4 AND 5).  
 FT CONFLICT 443 443 A -> V (IN REF. 4).  
 FT CONFLICT 552 552 L -> V (IN REF. 5).  
 FT CONFLICT 606 606 V -> M (IN REF. 2 AND 5).  
 FT CONFLICT 625 625 H -> Y (IN REF. 2 AND 5).  
 FT CONFLICT 625 625 Q -> H (IN REF. 2 AND 5).  
 FT SEQUENCE 708 AA; 76918 MW; E47BCF49FC78630A CRC64;  
 Query Match 20.3%; Score 70.5; DB 1; Length 708;  
 Best Local Similarity 32.8%; Pred. No. 8.9;  
 Matches 20; Conservative 7; Mismatches 21; Indels 13; Gaps 3;  
 QY 8 EERRPPTTTI-----PPSPPOKNQFGDDPLLGNSLSLETHPLYQSEPAVVPKTPPLKK 63  
 DB 413 EPTTPEDTAISLRPPSAPEMLTQH-----GAQEOLEDHPGHSSQAPIP--RADPLPR 463  
 QY 64 K 64  
 DB 464 R 464  
 RESULT 13  
 AD19\_MOUSE  
 ID AD19\_MOUSE STANDARD; PRT; 920 AA.  
 AC O35674;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DE ADAM 19 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase  
 DE domain 19) (Meltrin beta).

GN ADAM19 OR MLTNB.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Myoblasts;  
 RX MEDLINE=98129833; PubMed=9461614;  
 RA Inoue D., Reid M.S., Lum L., Kraetzschmar J., Westkamp G., Myung Y.M.,  
 RA Baron R., Blobel C.P.;  
 RT "Cloning and initial characterization of mouse meltrin beta and  
 RT analysis of the expression of four metalloprotease-disintegrins in  
 RT bone cells";  
 RL J. Biol. Chem. 273:4180-4187(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A., AND DEVELOPMENTAL STAGE.  
 RC TISSUE=Myoblasts;  
 RX MEDLINE=98286126; PubMed=9622634;  
 RA Kurisaki T., Masuda A., Osumi N., Nabeshima Y.-I., Fujisawa-Sehara A.;  
 RT "Spatially- and temporally-restricted expression of meltrin alpha  
 RT (ADAM12) and beta (ADAM19) in mouse embryo.";  
 RL Mech. Dev. 73:211-215(1998).  
 RN [3]  
 RP SEQUENCE OF 429-578 FROM N.A.  
 RC TISSUE=Embryonic fibroblast;  
 RX MEDLINE=96026308; PubMed=7566181;  
 RA Yagami-Hiromasa T., Sato T., Kurisaki T., Kamiyo K., Nabeshima Y.-I.,  
 RA Fujisawa-Sehara A.;  
 RT "A metalloprotease-disintegrin participating in myoblast fusion.";  
 RL Nature 377:652-656(1995).  
 RN [4]  
 RP FUNCTION.  
 RX MEDLINE=21153759; PubMed=11116142;  
 RA Shirakabe K., Wakatsuki S., Kurisaki T., Fujisawa-Sehara A.;  
 RT "Roles of Meltrin beta /ADAM19 in the processing of neuregulin.";  
 RL J. Biol. Chem. 276:9352-9358(2001).  
 CC -|- FUNCTION: PARTICIPATES IN THE PROTEOLYTIC PROCESSING OF BETA-TYPE  
 CC NEUREGULIN ISOFORMS WHICH ARE INVOLVED IN NEUROGENESIS AND  
 CC SYNAPTONEGNETIS, SUGGESTING A REGULATORY ROLE IN GLIAL CELL. ALSO  
 CC CLEAVES ALPHA-2 MACROGLOBULIN. MAY BE INVOLVED IN OSTEOBLAST  
 CC DIFFERENTIATION AND/OR OSTEOBLAST ACTIVITY IN BONE (BY  
 CC SIMILARITY).  
 CC -|- COFACTOR: BINDS ONE ZINC ION (BY SIMILARITY).  
 CC -|- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -|- TISSUE SPECIFICITY: WIDELY EXPRESSED, WITH THE HIGHEST EXPRESSION  
 CC IN BONE, HEART AND LUNG, FOLLOWED BY BRAIN AND SPLEEN AND  
 CC RELATIVELY LOW EXPRESSION IN LIVER, SKELETAL MUSCLE, KIDNEY AND  
 CC TESTIS. IN BONE, PRIMARILY EXPRESSED IN CELL OF THE OSTEOBLAST  
 CC LINEAGE AND NOT DETECTED IN MATURE OSTEOCLASTS.  
 CC -|- DEVELOPMENTAL STAGE: EXPRESSED IN THE HEART AND IN THE TAIL BUD AT  
 CC 8.0 DPC, AND THEN IN THE CRANIAL AND DORSAL ROOT GANGLIA. ALSO  
 CC EXPRESSED WEAKLY AND TRANSIENTLY IN THE INTESTINE, LUNG AND IN  
 CC BONE MARROW.  
 CC -|- INDUCTION: BY CALCITRIOL AND DURING OSTEOBLAST DIFFERENTIATION.  
 CC -|- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY  
 CC SIMILARITY).  
 CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.  
 CC -|- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.  
 CC -|- SIMILARITY: CONTAINS 1 DISINTEGRIN DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AF019887; AAC40037.1; -  
 CC EMBL; D50410; BAA18923.2; -  
 CC HSPSP; P18619; IFLV.  
 CC MEROPS; M12.214; -.

```
DR MGD: MGI:105377; Adam19.
DR InterPro: IPR001762; Disintegrin.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001818; Matrxin.
DR InterPro: IPR002870; Pep_M12B_propep.
DR InterPro: IPR001590; Reprolysin.
DR InterPro: IPR000130; Zn_MTPeptidse.
DR Pfam: PF00200; disintegrin; 1.
DR Pfam: PF01421; Reprolysin; 1.
DR Pfam: PF01562; Pep_M12B_propep; 1.
DR ProDom: PD000664; Disintegrin; 1.
DR SMART: SM00050; DISIN; 1.
DR PROSITE: PS02015; ADAM_MEPRO; 1.
DR PROSITE: PS00546; CYSTEINE_SWITCH; FALSE_NEG.
DR PROSITE: PS00427; DISINTEGRIN_1; FALSE_NEG.
DR PROSITE: PS02014; DISINTEGRIN_2; 1.
DR PROSITE: PS00022; EGF_1; FALSE_NEG.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Transmembrane; EGF-like domain; SH3-binding.
FT SIGNAL 1 26 POTENTIAL.
FT PROPEP 27 204 BY SIMILARITY.
FT CHAIN 205 920 ADAM 19.
FT DOMAIN 27 703 EXTRACELLULAR (POTENTIAL).
FT TRANSEM 704 724 POTENTIAL.
FT DOMAIN 725 920 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 211 409 METALLOPROTEASE.
FT DOMAIN 417 503 DISINTEGRIN-LIKE.
FT DOMAIN 435 438 POLY-GLU.
FT DOMAIN 504 653 CYS-RICH.
FT DOMAIN 654 686 EGF-LIKE.
FT DOMAIN 616 821 POLY-GLU.
FT SITE 835 841 SH3-BINDING (POTENTIAL).
FT SITE 840 846 SH3-BINDING (POTENTIAL).
FT SITE 133 133 CYSTEINE SWITCH (POTENTIAL).
FT METAL 346 346 ZINC (CATALYTIC).
FT ACT_SITE 347 347
FT METAL 350 350 ZINC (CATALYTIC).
FT METAL 356 356 ZINC (CATALYTIC).
FT DISULFID 321 404 BY SIMILARITY.
FT DISULFID 361 388 BY SIMILARITY.
FT DISULFID 466 488 BY SIMILARITY.
FT DISULFID 658 668 BY SIMILARITY.
FT DISULFID 662 674 BY SIMILARITY.
FT DISULFID 676 685 BY SIMILARITY.
FT CARBOHYD 145 145 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 445 445 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 649 649 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 920 AA; 100860 MW; 7094FDD4EE547382 CRC64;

Query Match 20.3%; Score 70.5; DB 1; Length 920;
Best Local Similarity 30.5%; Pred. No. 12;
Matches 25; Conservative 8; Mismatches 28; Indels 21; Gaps 4;

QY 4 RRYQERPPETTTTPPS-----SPQWKLO-----PGD--DPLIGNQSLE--- 42
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 829 RKESARRPPSRPPAPNCLLSQDFSRPPQKALPANVPVGQRTGRSGCTSLQPP 888
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 43 THPLYQSEPAVPVTKTPPLKK 64
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 889 SGQPQPRPPAVPVKPLPEYRSQ 910
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Query Match 20.1%; Score 70; DB 1; Length 785;
Best Local Similarity 33.3%; Pred. No. 11;
Matches 19; Conservative 5; Mismatches 23; Indels 10; Gaps 2;

QY 12 PETTTTPSP-----PQWKLPQGDPLLNQ-----SLLETHPLYQSEPAVPV 58
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 157 PYTAGVPEMSPVGTSVVQVTTATDADDDPTGNSARVVYSILQGGPYFSVEPKTG 213
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 15
SSP5_STRGN STANDARD; PRT; 1500 AA.
AC P16952; Q54184;
DT 01-AUG-1990 (Rel. 15, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Agglutinin receptor precursor (SSP-5).
```

```
GN CDH7.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=White leghorn; TISSUE=Brain;
RX MEDLINE=95309115; PubMed=7540531;
RA Nakagawa S., Takeichi M.;
RT "Neural crest cell-cell adhesion controlled by sequential and
RT subpopulation-specific expression of novel cadherins.";
RL Development 121:1321-1332(1995).
CC -|- FUNCTION: CADHERINS ARE CALCIIUM DEPENDENT CELL ADHESION PROTEINS.
CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC SORTING OF HETEROGENEOUS CELL TYPES.
CC -|- SUBCELLULAR LOCATION: Type I membrane protein.
CC -|- SIMILARITY: CONTAINS 5 CADHERIN DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D42150; BAA07721.1; -.
DR HSSP: P15116; INCH.
DR InterPro: IPR002126; Cadherin.
DR InterPro: IPR000233; Cadherin_C_term.
DR Pfam: PF00028; cadherin; 5.
DR Pfam: PF01049; Cadherin_C_term; 1.
DR PRINTS: PR00205; CADHERIN.
DR SMART: SM00112; CA; 5.
DR PROSITE: PS00232; CADHERIN_1; 3.
DR PROSITE: PS0268; CADHERIN_2; 5.
KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
KW Signal.
FT SIGNAL 1 27 POTENTIAL.
FT PROPEP 28 47 POTENTIAL.
FT CHAIN 48 785 CADHERIN-7.
FT DOMAIN 48 607 EXTRACELLULAR (POTENTIAL).
FT TRANSEM 608 628 POTENTIAL.
FT DOMAIN 629 785 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 49 153 CADHERIN 1.
FT DOMAIN 154 262 CADHERIN 2.
FT DOMAIN 263 377 CADHERIN 3.
FT DOMAIN 378 482 CADHERIN 4.
FT DOMAIN 482 599 CADHERIN 5.
FT CARBOHYD 449 449 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 530 530 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 785 AA; 8171 MW; 895B06D8141E34D4 CRC64;

Query Match 20.1%; Score 70; DB 1; Length 785;
Best Local Similarity 33.3%; Pred. No. 11;
Matches 19; Conservative 5; Mismatches 23; Indels 10; Gaps 2;

QY 12 PETTTTPSP-----PQWKLPQGDPLLNQ-----SLLETHPLYQSEPAVPV 58
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 157 PYTAGVPEMSPVGTSVVQVTTATDADDDPTGNSARVVYSILQGGPYFSVEPKTG 213
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 15
SSP5_STRGN STANDARD; PRT; 1500 AA.
AC P16952; Q54184;
DT 01-AUG-1990 (Rel. 15, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Agglutinin receptor precursor (SSP-5).
```



GN SSP5 OR SSPB.  
OS Streptococcus gordonii.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1302;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=M5;  
RX MEDLINE=90236997; PubMed=2185241;  
RA Demuth D.R., Golub E.E., Malamud D.;  
RT "Streptococcal-host interactions. Structural and functional analysis  
of a Streptococcus sanguis receptor for a human salivary  
glycoprotein.";  
RN J. Biol. Chem. 265:7120-7126(1990).  
RN [2]  
RP REVISIONS, SEQUENCE FROM N.A.  
RC STRAIN=M5;  
RX MEDLINE=96310377; PubMed=8733238;  
RA Demuth D.R., Duan Y., Brooks W., Holmes A.R., McNab R.,  
Jenkinson H.F.;  
RT "Tandem genes encode cell-surface polypeptides SspA and SspB which  
mediate adhesion of the oral bacterium Streptococcus gordonii to  
human and bacterial receptors";  
RL Mol. Microbiol. 20:403-413(1996).  
CC -!- FUNCTION: MAY BIND SIALIC ACID RESIDUES OF SALIVARY AGGLUTININ  
(SAG) IN A CALCIUM-DEPENDENT REACTION. THE INTERACTION OF SAG WITH  
ITS RECEPTOR IN VARIOUS ORAL STREPTOCOCCI MODULATE BACTERIAL  
COLONIZATION OF ORAL TISSUE AND IS ASSOCIATED WITH REDUCED LEVELS  
OF DENTAL CARIES.  
CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by  
an amide bond (Potential).  
CC -!- SIMILARITY: BELONGS TO THE SPAP/SSP5/SPAA FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; UA0026; AAC44100.1; -.  
DR PIR; A35186; A35186.  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
DR ProDom; PD153432; Csurface\_antigen; 2.  
DR TIGRFAMs; TIGR01167; LPXTG\_anchor; 1.  
DR PROSITE; PS50847; GRAM\_POS\_ANCHORING; 1.  
Cell wall; Peptidoglycan-anchor; Calcium-binding; Repeat; Signal.  
SIGNAL  
1 38  
FT CHAIN 39 1470 AGGLUTININ RECEPTOR.  
FT PROPEP 1471 1500 REMOVED BY SORTASE (POTENTIAL).  
FT SIMILAR 164 470 TO M PROTEIN OF S. PYOGENES.  
FT DOMAIN 164 470 4 X APPROXIMATE TANDEM REPEATS, HRL.  
FT REPEAT 164 241 1.  
FT REPEAT 242 323 2.  
FT REPEAT 324 405 3.  
FT REPEAT 406 470 4.  
FT DOMAIN 771 887 3 X APPROXIMATE TANDEM REPEATS, PRI.  
FT DOMAIN 1414 1436 PRO-RICH (PR2).  
FT CA\_BIND 220 235 POTENTIAL.  
FT CA\_BIND 301 316 POTENTIAL.  
FT CA\_BIND 931 950 POTENTIAL.  
FT CA\_BIND 1300 1315 POTENTIAL.  
FT SITE 1467 1471 LPXTG SORTING SIGNAL (POTENTIAL).  
FT MOD\_RES 1470 1470 AMIDE-LINKED TO CELL WALL (POTENTIAL).  
SQ SEQUENCE 1500 AA; 164552 MW; DCFI90E7D4D889F CRC64;  
  
Query Match 19.8%; Score 69; DB 1; Length 1500;  
Best Local Similarity 29.9%; Pred. No. 28;  
Matches 20; Conservative 11; Mismatches 22; Indels 14; Gaps 3;  
  
QY 6 YQERPPETTTTPPS-----SPQWKLPQGGDDPLGNQSLLETHPL---YQSEPA 52

Db 771 YETEKPLEPAPVAPSYENEPTPPVKTPDQPEPSKPEEPTYETEKPLEPAPVAPSYENEPT 830  
Qy 53 VPVIKTP 59  
Db 831 PPV-KTP 836

Search completed: January 2, 2003, 13:44:53  
Job time : 6.16129 secs





AC Q66399;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DE Surface antigen.  
OS Duck hepatitis B virus (DHBV).  
OC Viruses; Retroid viruses; Hepadnaviridae; Avihepadnavirus.  
OX NCBI\_TaxID=12639;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=INDIAN;  
RA Munshi A., Panda S.K.;  
RT "Cloning sequencing and sequence comparison of the Indian isolate.";  
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; X74623; CAA52699.1; -  
DR InterPro; IPR000349; Hepvir\_surfa.  
DR Pfam; PF00695; VMSA; 1.  
SQ SEQUENCE 364 AA; 40385 MW; E2E27FB4E4775C19 CRC64;  
  
Query Match 100.0%; Score 348; DB 12; Length 364;  
Best Local Similarity 100.0%; Pred. No. 9.4e-32;  
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 EAFRYQERPPETTTPSSPPQKLOPGDDPLLGNSLLETHLYQSEPAVPVKTTP 60  
Db 134 EAFRYQERPPETTTPSSPPQKLOPGDDPLLGNSLLETHLYQSEPAVPVKTTP 193  
  
QY 61 LKKK 64  
Db 194 LKKK 197  
  
RESULT 3  
Q80QX1 PRELIMINARY; PRT; 328 AA.  
AC Q80QX1;  
DT 01-JUN-2002 (TREMBlrel. 21, Created)  
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
DE Presurface protein.  
OS Duck hepatitis B virus (DHBV).  
OC Viruses; Retroid viruses; Hepadnaviridae; Avihepadnavirus.  
OX NCBI\_TaxID=12639;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=INDIANA;  
RA Wang C.-Y.J., Giambone J.J., Dormitorio T.V.;  
RT "The complete sequence of Duck Hepatitis B virus Indiana isolate.";  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF493986; AAM11781.1; -  
SQ SEQUENCE 328 AA; 36146 MW; B6D1C9E73FA1556 CRC64;  
  
Query Match 96.0%; Score 334; DB 12; Length 328;  
Best Local Similarity 96.9%; Pred. No. 3.3e-30;  
Matches 62; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 EAFRYQERPPETTTPSSPPQKLOPGDDPLLGNSLLETHLYQSEPAVPVKTTP 60  
Db 98 EAFRYQERPPETTTPSSPPQKLOPGDDPLLGNSLLETHLYQSEPAVPVKTTP 157  
  
QY 61 LKKK 64  
Db 158 LKKK 161  
  
RESULT 4  
O72885 PRELIMINARY; PRT; 330 AA.  
AC O72885;  
DT 01-AUG-1998 (TREMBlrel. 07, Created)  
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
DE Surface protein.

OS Duck hepatitis B virus (DHBV).  
OC Viruses; Retroid viruses; Hepadnaviridae; Avihepadnavirus.  
OX NCBI\_TaxID=12639;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AUSTRALIAN DHBV;  
RA Triyatni M.;  
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AUSTRALIAN DHBV;  
RA MEDLINE=21102973; PubMed=11161276;  
RT Triyatni M., Ey P.L., Tran T., Le Mire M., Qiao M., Burrell C.J.,  
RL Jilbert A.R.;  
RT "Sequence comparison of an Australian duck hepatitis B virus strain  
with other avian hepadnaviruses";  
RL J. Gen. Virol. 82:373-378(2001).  
DR EMBL; AJ006350; CAA06988.1; -  
DR InterPro; IPR000349; Hepvir\_surfa.  
DR Pfam; PF00695; VMSA; 1.  
SQ SEQUENCE 330 AA; 36789 MW; EDED4F42373ADA99 CRC64;  
  
Query Match 87.8%; Score 305.5; DB 12; Length 330;  
Best Local Similarity 89.2%; Pred. No. 6e-27;  
Matches 58; Conservative 3; Mismatches 3; Indels 1; Gaps 1;  
  
QY 1 EAFRYQERPPETTTPSSPPQKLOPGDDPLLGNSLLETHLYQSEPAVPVKTTP 59  
Db 98 EAFRYQERPPETTTPSSPPQKLOPGDDPLLGNSLLETHLYQSEPAVPVKTTP 157  
  
QY 60 LKKK 64  
Db 158 LKKK 162  
  
RESULT 5  
Q66405 PRELIMINARY; PRT; 330 AA.  
AC Q66405;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Pre-S protein.  
OS Duck hepatitis B virus (DHBV).  
OC Viruses; Retroid viruses; Hepadnaviridae; Avihepadnavirus.  
OX NCBI\_TaxID=12639;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=DHBVQCA34;  
RA Tong S., Mattes F., Blum H.E., Fernholz D., Schneider R., Will H.;  
RT "Complete nucleotide sequence of a Chinese Hepatitis B virus";  
RL Submitted (JUN-1991) to the EMBL/GenBank/DBJ databases.  
DR EMBL; X60213; CAA42771.1; -  
DR InterPro; IPR000349; Hepvir\_surfa.  
DR Pfam; PF00695; VMSA; 1.  
SQ SEQUENCE 330 AA; 36959 MW; 7CE142013BB8D9F4 CRC64;  
  
Query Match 85.5%; Score 297.5; DB 12; Length 330;  
Best Local Similarity 87.7%; Pred. No. 4.9e-26;  
Matches 57; Conservative 3; Mismatches 4; Indels 1; Gaps 1;  
  
QY 1 EAFRYQERPPETTTPSSPPQKLOPGDDPLLGNSLLETHLYQSEPAVPVKTTP 59  
Db 98 EAFRYQERPPETTTPSSPPQKLOPGDDPLLGNSLLETHLYQSEPAVPVKTTP 157  
  
QY 60 LKKK 64  
Db 158 LKKK 162  
  
RESULT 6  
Q66404 PRELIMINARY; PRT; 366 AA.  
ID Q66404

```
AC Q06404; 1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Pre-5/S.
OS Duck hepatitis B virus (DBHV).
OC Viruses; Retroid viruses; Hepadnaviridae; Avihepadnavirus.
OX NCBI_TaxID=12639;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DRBVCA34;
RA Tong S., Mattes F., Blum H.E., Fernholz D., Schneider R., Will H.;
RT "Complete nucleotide sequence of a Chinese Hepatitis B virus."
RL Submitted (JUN-1991) to the EMBL/GenBank/DBJ databases.
DR EMBL; X60213; CAA42770.1; -
DR InterPro; IPR000349; Hepvir_surfa.
DR Pfam; PF00695; VMSA; 1.
SQ SEQUENCE 366 AA; 41057 MW; FDF3616EBC39629D CRC64;

Query Match      85.5%; Score 297.5; DB 12; Length 366;
Best Local Similarity 87.7%; Pred. No. 5.5e-26;
Matches 57; Conservative 3; Mismatches 4; Indels 1; Gaps 1;

QY 1 EAFRYOERPPETTTTPSPQWKLQGGDDPLLGNSLLETHPLXQS-EPAPVVIKTP 59
   |||||
DB 134 EAFRYOERPPETTTTAPTPTPWKLQGGDDPLLEKNKSLLETHPLXQNPPEPAPVVIKTP 193
   |||||

QY 60 PLKKK 64
   |||||
DB 194 PLKKK 198

RESULT 7
Q91HP5
ID Q91HP5 PRELIMINARY; PRT; 329 AA.
AC Q91HP5;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE PreS protein.
OS Duck hepatitis B virus (DBHV).
OC Viruses; Retroid viruses; Hepadnaviridae; Avihepadnavirus.
OX NCBI_TaxID=12639;
RN [1]
RP SEQUENCE FROM N.A.
RA Hu J., Tang N., Huang A.;
RT "Sequence Analysis of a Cloned Duck Hepatitis B Virus Genome from
   Chongqing Brown Duck."
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF404406; AAK85437.1; -
DR InterPro; IPR000349; Hepvir_surfa.
DR Pfam; PF00695; VMSA; 1.
SQ SEQUENCE 329 AA; 36436 MW; F1DFE48192CE9F97 CRC64;

Query Match      84.1%; Score 292.5; DB 12; Length 329;
Best Local Similarity 86.2%; Pred. No. 1.8e-25;
Matches 56; Conservative 3; Mismatches 5; Indels 1; Gaps 1;

QY 1 EAFRYOERPPETTTTPSPQWKLQGGDDPLLGNSLLETHPLXQS-EPAPVVIKTP 59
   |||||
DB 98 EAFRYOERPPETTTTAPTPTPWKLQGGDDPLLEKNKSLLETHPLXQNPPEPAPVVIKTP 157
   |||||

QY 60 PLKKK 64
   |||||
DB 158 PLKKK 162

RESULT 8
Q9WFA3
ID Q9WFA3 PRELIMINARY; PRT; 329 AA.
AC Q9WFA3;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
```

```
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE PreS antigen.
OS snow goose hepatitis B virus.
OC Viruses; Retroid viruses; Hepadnaviridae; Avihepadnavirus.
OX NCBI_TaxID=89623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SGHBV1-13;
RX MEDLINE=99420377; PubMed=10489339;
RA Chang S.-F., Netter H.J., Bruns M., Schneider R., Froelich K., Will H.;
RT "A new avian hepadnavirus infecting snow geese (Anser caerulescens)
   produces a significant fraction of virions containing single-stranded
   DNA."
RL Virology 262:39-54(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SGHBV1-13;
RA Chang S.-F., Netter H.J., Bruns M., Schneider R., Froelich K.,
   Will H.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF110996; AAD21982.1; -
DR InterPro; IPR000349; Hepvir_surfa.
DR Pfam; PF00695; VMSA; 1.
SQ SEQUENCE 329 AA; 36657 MW; 12B2DC5E8B7FC420 CRC64;

Query Match      69.8%; Score 243; DB 12; Length 329;
Best Local Similarity 78.5%; Pred. No. 8e-20;
Matches 51; Conservative 3; Mismatches 9; Indels 2; Gaps 2;

QY 1 EAFRYOERPPETTTTPSPQWKLQGGDDPLLGNSLLETHPLXQ-SEPAPVVIKTP 59
   |||||
DB 98 EAFRYOERPPETTTTTPP-TPTPWKLQGGDDPLLGNSLLETHPLXQNPPEPAPVVIKTP 156
   |||||

QY 60 PLKKK 64
   |||||
DB 157 LVKKK 161

RESULT 9
Q9WFA9
ID Q9WFA9 PRELIMINARY; PRT; 329 AA.
AC Q9WFA9;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE PreS antigen.
OS snow goose hepatitis B virus.
OC Viruses; Retroid viruses; Hepadnaviridae; Avihepadnavirus.
OX NCBI_TaxID=89623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SGHBV1-19;
RX MEDLINE=99420377; PubMed=10489339;
RA Chang S.-F., Netter H.J., Bruns M., Schneider R., Froelich K., Will H.;
RT "A new avian hepadnavirus infecting snow geese (Anser caerulescens)
   produces a significant fraction of virions containing single-stranded
   DNA."
RL Virology 262:39-54(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SGHBV1-19;
RA Chang S.-F., Netter H.J., Bruns M., Schneider R., Froelich K.,
   Will H.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF110998; AAD21992.1; -
DR InterPro; IPR000349; Hepvir_surfa.
DR Pfam; PF00695; VMSA; 1.
SQ SEQUENCE 329 AA; 36645 MW; D896E4B2979F201E CRC64;

Query Match      69.8%; Score 243; DB 12; Length 329;
Best Local Similarity 78.5%; Pred. No. 8e-20;
Matches 51; Conservative 3; Mismatches 9; Indels 2; Gaps 2;
```





**This Page Blank (uspto)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2003, 13:39:28 ; Search time 8.25403 Seconds  
(without alignments)  
371.305 Million cell updates/sec

Title: US-09-818-066-34\_COPY\_80\_102

Perfect score: 128

Sequence: 1 PQEIPQWTPEDQAREAFRR 23

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	128	100.0	161	AAW11854	Duck hepatitis B v
2	54	42.2	1115	AAW11854	Putative P. abyssi
3	53	41.4	485	22	Peptide #3828 enco
4	53	41.4	485	22	Peptide #3880 enco
5	53	41.4	485	22	Protein #3727 enco
6	53	41.4	485	22	Human brain expres
7	53	41.4	485	22	Human bone marrow
8	53	41.4	485	22	Peptide #3795 enco
9	53	41.4	485	22	Peptide #3908 enco
10	53	41.4	485	22	Peptide #3733 enco

11	53	41.4	485	23	ABG39157	Human peptide enco
12	53	41.4	536	22	ABB50227	Human transcriptio
13	53	41.4	545	23	ABB08528	Human zinc finger
14	51.5	40.2	51	22	AAU44633	Propionibacterium
15	51	39.8	150	21	AAG27027	Zea mays protein f
16	50	39.1	172	19	AAW64201	Human calcitriactin
17	50	39.1	172	21	AAG03863	Human secreted pro
18	50	39.1	185	23	ABP41802	Human ovarian anti
19	50	39.1	512	22	ABB57841	Drosophila melanog
20	50	39.1	528	20	AAV05909	Thermophilus therm
21	50	39.1	529	19	AAW85044	Tau subunit of a D
22	50	39.1	529	20	AAV50003	Thermus thermophil
23	50	39.1	529	20	AAV50025	Thermus thermophil
24	50	39.1	1824	21	AAV49431	Murine CACNAIF pro
25	49	38.3	26	22	ABB38481	Peptide #5987 enco
26	49	38.3	26	22	ABB23637	Protein #5636 enco
27	49	38.3	26	22	AAW59094	Human brain expres
28	49	38.3	26	22	AAW71629	Human bone marrow
29	49	38.3	26	22	AAW31923	Peptide #5960 enco
30	49	38.3	26	23	ABG41443	Human peptide enco
31	49	38.3	45	21	AAW33398	Pinus radiata tran
32	49	38.3	61	21	AAW33369	Pinus radiata tran
33	49	38.3	62	21	AAW33404	Pinus radiata tran
34	49	38.3	66	21	AAW33299	Pinus radiata tran
35	49	38.3	78	21	AAW33205	Pinus radiata tran
36	49	38.3	146	21	AAW33155	Euclalyptus grandis
37	49	38.3	156	21	ABG44280	Arabidopsis thalia
38	49	38.3	216	23	ABP30974	Protein encoded by
39	49	38.3	324	22	AAE01923	Arabidopsis thalia
40	49	38.3	326	23	AAU93118	Arabidopsis transc
41	49	38.3	345	21	AAW34523	Arabidopsis thalia
42	49	38.3	352	23	AAU93021	Arabidopsis transc
43	49	38.3	365	21	AAW30853	Arabidopsis thalia
44	49	38.3	412	21	AAW33303	Pinus radiata tran
45	49	38.3	469	23	AAU71802	MYB factor-1 (MYB-

#### ALIGNMENTS

RESULT 1

AAW11854  
ID AAW11854 standard; Protein; 161 AA.

AC AAW11854;

DT 10-MAY-1997 (first entry)

DE Duck hepatitis B virus pre-S domain.

KW Hepadnavirus receptor; p120; p170; vaccine; pre-S domain.

XX Duck hepatitis B virus.

PH Key Location/Qualifiers

FT Binding-site 87..102

FT /label= p170\_binding\_site

FT /note= "the p170 binding site has been mapped to a major neutralising epitope of the pre-S domain (aa87-102), within which the Lys-95 and Arg-97 residues required for virion-receptor interaction".

FT Peptide 1..102

FT /note= "claimed pre-S fragment (Claim 20)

FT capable of binding hepadnavirus receptor"

FT Peptide 1..104

FT /note= "claimed pre-S fragment (Claim 20)

FT capable of binding hepadnavirus receptor"

FT Peptide 1..126

FT /note= "claimed pre-S fragment (Claim 20)

FT capable of binding hepadnavirus receptor"

FT Peptide 1..138

FT /note= "claimed pre-S fragment (Claim 20)

FT Peptide 25..102 capable of binding hepadnavirus receptor"  
FT /note= "claimed pre-S fragment (Claim 20)  
FT capable of binding hepadnavirus receptor"  
FT Peptide 25..104  
FT /note= "claimed pre-S fragment (Claim 20)  
FT capable of binding hepadnavirus receptor"  
FT Peptide 25..126  
FT /note= "claimed pre-S fragment (Claim 20)  
FT capable of binding hepadnavirus receptor"  
FT Peptide 25..161  
FT /note= "claimed pre-S fragment (Claim 20)  
FT capable of binding hepadnavirus receptor"  
FT Peptide 42..102  
FT /note= "claimed pre-S fragment (Claim 20)  
FT capable of binding hepadnavirus receptor"  
FT Peptide 59..104  
FT /note= "claimed pre-S fragment (Claim 20)  
FT capable of binding hepadnavirus receptor"  
FT Peptide 59..126  
FT /note= "claimed pre-S fragment (Claim 20)  
FT capable of binding hepadnavirus receptor"  
FT Peptide 59..161  
FT /note= "claimed pre-S fragment (Claim 20)  
FT capable of binding hepadnavirus receptor"  
FT Peptide 71..126  
FT /note= "claimed pre-S fragment (Claim 20)  
FT capable of binding hepadnavirus receptor"  
FT Peptide 80..104  
FT /note= "claimed pre-S fragment (Claim 20)  
FT capable of binding hepadnavirus receptor"  
FT Peptide 80..161  
FT /note= "claimed pre-S fragment (Claim 20)  
FT capable of binding hepadnavirus receptor"  
FT Peptide 87..161  
FT /note= "claimed pre-S fragment (Claim 20)  
FT capable of binding hepadnavirus receptor"  
FT Peptide 87..104  
FT /note= "claimed pre-S fragment (Claim 21)  
FT capable of binding hepadnavirus receptor"  
FT XX  
PN WO9704000-A1.  
XX  
XX PD 06-FEB-1997.  
XX  
XX PD 22-JUL-1996; 96WO-US12098.  
XX  
XX PD 21-JUL-1995; 95US-0001371.  
XX  
XX PA (GEO) GEN HOSPITAL CORP.  
XX  
XX PL Li J, Tong S, Wands JR;  
XX  
XX WPI; 1997-132572/12.  
DR N-PSDB; AAT59583.  
XX  
XX PT p170, a new avian hepadnavirus receptor - binds to pre-S domain of  
PT duck hepatitis B virus, attenuated p170 may be used to immunise  
PT animals against hepadnaviral infection  
XX  
XX PS Claim 19; Page 87-88; 175pp; English.  
XX  
XX CC The pre-S domain (AAW11854) of the duck hepatitis B virus (DHBV)  
CC large envelope protein binds to the cellular p170 receptor (see  
CC also AAW11851-52) at a major neutralising epitope, within which are 2  
CC basic amino acids (Lys-95, Arg-97) required for virion-receptor  
CC interaction. Peptide sequences encompassing Lys-95 and Arg-97  
CC can act as p170 ligands. Such peptides can be used to reduce  
CC the level of hepadnaviral infection in an animal. The pre-S  
CC domain (esp. attenuated), and nucleic acids encoding it, can also

CC be used as vaccines to immunise animals against hepadnavirus  
CC infection.  
XX  
SQ Sequence 161 AA;  
Query Match 100.0%; Score 128; DB 18; Length 161;  
Best Local Similarity 100.0%; Pred. No. 4.4e-09;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 PQEIPQPTPEEDQKAREAFRR 23  
DB 80 PQEIPQPTPEEDQKAREAFRR 102  
RESULT 2  
AAB96812  
ID AAB96812 standard; Protein; 1115 AA.  
XX  
AC AAB96812;  
XX  
DT 29-OCT-2001 (first entry)  
XX  
DE Putative P. abyssi cell division control protein.  
XX  
KW Hyperthermophilic archaeon; hyperthermophilic protein.  
XX  
OS Pyrococcus abyssi.  
XX  
PN FR2792651-A1.  
XX  
PD 27-OCT-2000.  
XX  
PF 21-APR-1999; 99FR-0005034.  
XX  
PR 21-APR-1999; 99FR-0005034.  
XX  
PA (CNRS) CNRS CENT NAT RECH SCI.  
PA (IFRE-) IFREMER INST FR RECH EXPL MER.  
XX  
PI Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;  
PI Querellou J, Weissenbach J, Saurin W, Heilig R;  
XX  
DR WPI; 2001-126236/14.  
XX  
XX New nucleotide sequences isolated from Pyrococcus abyssi encode  
XX proteins useful in industry -  
PS Claim 7; Pages 1589-1592; 1657pp; French.  
XX  
CC The present invention relates to the genomic sequence of Pyrococcus  
CC abyssi (see AAF86431 and AAF41223-7) and P. abyssi proteins. P. abyssi is  
CC a hyperthermophilic archaeon, which is isolated from deep-sea  
CC hydrothermal vents. The present sequence is one such P. abyssi protein.  
CC The proteins of the present invention have various potential industrial  
CC uses, since the proteins are stable at very high temperatures, some up to  
CC 110 degrees centigrade.  
CC Note: This patent is in the same patent family as WO2000065062, which  
CC contains additional sequences as shown in AAB99132-AAB99143,  
CC AAF75903-AAH75920 and AAG66436.  
XX  
SQ Sequence 1115 AA;  
Query Match 42.2%; Score 54; DB 22; Length 1115;  
Best Local Similarity 50.0%; Pred. No. 1.4e+02;  
Matches 11; Conservative 5; Mismatches 6; Indels 0; Gaps 0;  
Qy 2 QEIPOQPTPEEDQKAREAFRR 23  
DB 257 KEIELEITPEEQKIRELAKR 278  
RESULT 3  
ABB31177





```

PN WO200157274-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00666.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488999/53.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
XX hearts -
XX
PS Claim 15; SEQ ID NO 23498; 530pp; English.
XX
CC The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart (see
CC ABA21535-ABA41305). The present sequence is a protein encoded by one such
CC probe. The probes may be used for predicting, measuring and displaying
CC gene expression in samples derived from the human heart via microarrays.
CC By measuring gene expression, the probes are useful for predicting,
CC diagnosing, grading, staging, monitoring and prognosing diseases of the
CC human heart and vascular system e.g. cardiovascular disease,
CC hypertension, cardiac arrhythmias and congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 485 AA;
XX
Query Match 41.4%; Score 53; DB 22; Length 485;
Best Local Similarity 50.0%; Pred. No. 79;
Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
QY 5 POPQWTPPEEDQKAREAFR 22
DB 126 PQSEWKPYRDTEDEAFQ 143
II : | | | : |||:
RESULT 6
AAM57139
ID AAM57139 standard; Protein; 485 AA.
XX
AC AAM57139;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe encoded protein SEQ ID NO: 29244.
XX
KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer.
XX
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00667.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
XX
PT Human genome-derived single exon nucleic acid probes useful for
analyzing gene expression in human bone marrow -

```

---

```

PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
XX brains -
XX
PS Example 4; SEQ ID NO: 29244; 650pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is a protein encoded by one of
CC the probes of the invention.
XX
SQ Sequence 485 AA;
XX
Query Match 41.4%; Score 53; DB 22; Length 485;
Best Local Similarity 50.0%; Pred. No. 79;
Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
QY 5 POPQWTPPEEDQKAREAFR 22
DB 126 PQSEWKPYRDTEDEAFQ 143
II : | | | : |||:
RESULT 7
AAM69532
ID AAM69532 standard; Protein; 485 AA.
XX
AC AAM69532;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed probe encoded protein SEQ ID NO: 29838.
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00668.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
analyzing gene expression in human bone marrow -

```

XX PS Example 4; SEQ ID NO: 29838; 658pp + Sequence Listing; English.

XX CC The present invention provides a number of single exon nucleic acid

CC probes which are derived from genomic sequences expressed in the human

CC bone marrow. They can be used to measure gene expression in bone marrow

CC samples, which may enable the improved diagnosis and treatment of cancers

CC such as lymphoma, leukemia and myeloma. The present sequence is a

CC protein encoded by one of the probes of the invention.

XX SQ Sequence 485 AA;

Query Match 41.4%; Score 53; DB 22; Length 485;

Best Local Similarity 50.0%; Pred. No. 79;

Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 5 PQPQWTPEDQKAREAFR 22

Db 126 PQSEWKPYRDTEDREAFQ 143

II : I I I : I I I I :

RESULT 8

AAM17361

ID AAM17361 standard; Protein; 485 AA.

XX AC AAM17361;

XX DT 12-OCT-2001 (first entry)

XX DE Peptide #3795 encoded by probe for measuring cervical gene expression.

XX KW Probe; human; microarray; gene expression; cervical epithelial cell;

XX KW cervical cancer.

XX OS Homo sapiens.

XX PN WO200157278-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00670.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2001-488901/53.

XX PT Human genome-derived single exon nucleic acid probes useful for

XX PT analyzing gene expression in human cervical epithelial cells -

XX PS Claim 27; SEQ ID NO 22187; 487pp; English.

XX CC The present invention relates to human single exon nucleic acid probes

CC (SNP: see AAI10068-AAI28459). The present sequence is a peptide encoded

CC by one such probe. The SNPs are derived from human HeLa cells. The SNPs

CC can be used to produce a single exon microarray, which can be used for

CC measuring human gene expression in a sample derived from human cervical

CC epithelial cells. By measuring gene expression, the probes are therefore

CC useful in grading and/or staging of diseases of the cervix, notably

CC cervical cancer.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX

XX SQ Sequence 485 AA;

Query Match 41.4%; Score 53; DB 22; Length 485;

Best Local Similarity 50.0%; Pred. No. 79;

Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 5 PQPQWTPEDQKAREAFR 22

Db 126 PQSEWKPYRDTEDREAFQ 143

II : I I I : I I I I :

RESULT 9

AAM29871

ID AAM29871 standard; Protein; 485 AA.

XX AC AAM29871;

XX DT 17-OCT-2001 (first entry)

XX DE Peptide #3908 encoded by probe for measuring placental gene expression.

XX KW Probe; microarray; human; placenta; antenatal diagnosis;

XX KW genetic disorder.

XX OS Homo sapiens.

XX PN WO200157272-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00663.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2001-488897/53.

XX PT Human genome-derived single exon nucleic acid probes useful for

XX PT analyzing gene expression in human placenta -

XX PS Claim 27; SEQ ID NO 30140; 654pp; English.

XX CC The present invention relates to single exon nucleic acid probes (SNP:

CC see AAI13115-AAI57546). The present sequence is a peptide encoded by one

CC such probe. The probes are useful for producing a microarray for

CC predicting, measuring and displaying gene expression in samples derived

CC from human placenta. The probes are useful for antenatal diagnosis of

CC human genetic disorders.

XX SQ Sequence 485 AA;

Query Match 41.4%; Score 53; DB 22; Length 485;

Best Local Similarity 50.0%; Pred. No. 79;

Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 5 PQPQWTPEDQKAREAFR 22

Db 126 PQSEWKPYRDTEDREAFQ 143

II : I I I : I I I I :

RESULT 10

AAM05051

ID AAM05051 standard; Protein; 485 AA.

XX



CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 485 AA;

Query Match 41.4%; Score 53; DB 23; Length 485;  
Best Local Similarity 50.0%; Pred. No. 79;  
Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 5 PQQWTPEDQKAREAFR 22  
||:|:|:|:|:  
Db 126 PQSEKPYRDTEDREAFQ 143

RESULT 12  
ABB50227  
ID ABB50227 standard; Protein; 536 AA.

XX  
XX  
AC ABB50227;  
XX  
XX 05-FEB-2002 (first entry)  
XX Human transcription factor TRFX-78.

XX Human; transcription factor; TRFX; cell proliferative disease;  
XX autoimmune disease; inflammation; neurological disease;  
XX developmental disorder; cancer; AIDS; infection; cytostatic; anti-HIV;  
XX neuroprotective; antiinflammatory; gene therapy.

XX Homo sapiens.

XX WO200172777-A2.

XX 04-OCT-2001.

XX 13-MAR-2001; 2001WO-US08117.

XX 13-MAR-2000; 2000US-0188986.

XX (INCY-) INCYTE GENOMICS INC.

XX Hillman JL, Baughn MR, Yue H, Lal P, Lu DAM, Patterson C;  
XX Atiznola Y, Bandman O, Tang YT, Mathur P, Shah P, Au-Young J;  
XX Reddy R;

XX WPI; 2001-570896/64.

XX N-PSDB; ABA83051.

XX Novel transcription factor polypeptides, used to treat diseases  
XX associated with altered activity and expression of TRFX, and to screen  
XX for agents capable of modulating its activity -

XX Claim 1; Pages 220-221; 327pp; English.

XX The present sequence is the protein sequence for a human transcription  
XX factor. The transcription factor and its coding sequence are useful in  
XX the diagnosis, treatment and prevention of diseases associated with  
XX altered expression of the transcription factor e.g. cell proliferative,  
XX autoimmune/inflammatory, neurological and developmental disorders. A  
XX number of specific disorders/diseases are given in the specification,  
XX including: arteriosclerosis, cirrhosis, hepatitis, cancers, AIDS,  
XX allergies, anaemia, asthma, autoimmune thyroiditis, bronchitis, atopic  
XX dermatitis, diabetes mellitus, emphysema, Goodpasture's syndrome, gout,  
XX Grave's disease, multiple sclerosis, osteoarthritis, pancreatitis,  
XX psoriasis, rheumatoid arthritis, systemic lupus erythematosus, ulcerative  
XX colitis, uveitis, Alzheimer's disease, Huntington's disease, Parkinson's  
XX disease, stroke, and viral, bacterial, fungal and protozoal infections.

XX Sequence 536 AA;

Query Match 41.4%; Score 53; DB 22; Length 536;

Best Local Similarity 50.0%; Pred. No. 88;

Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 5 PQQWTPEDQKAREAFR 22  
||:|:|:|:|:  
Db 177 PQSEKPYRDTEDREAFQ 194

RESULT 13  
ABB08528  
ID ABB08528 standard; Protein; 545 AA.

XX  
XX ABB08528;

XX 01-JUL-2002 (first entry)

XX Human zinc finger protein 60.

XX Zinc finger; human; tumour; HIV; cytostatic; anti-HIV.

XX Homo sapiens.

XX CN1321688-A.

XX 14-NOV-2001.

XX 29-APR-2000; 2000CN-0115576.

XX 29-APR-2000; 2000CN-0115576.

XX (BODE-) BODE GENE DEV CO LTD SHANGHAI.

XX Mao Y, Xie Y;

XX WPI; 2002-140643/19.

XX N-PSDB; ABA99168.

XX Novel polypeptide-human zinc finger protein 60 and polynucleotide for  
XX coding this polypeptide -

XX Claim 1; Page 27-28(disclosure); 35pp; Chinese.

XX This invention relates to a novel polypeptide-human zinc finger protein  
XX 60, polynucleotide for coding this polypeptide and method for producing  
XX said polypeptide by using DNA recombination technology. The invention  
XX also discloses the method for curing several diseases, such as various  
XX tumours, diseases of nervous system, development disturbance, some  
XX genetic diseases, diseases of endocrine system and HIV infection  
XX using the said polypeptide. The invention also discloses an antagonist  
XX for resisting said polypeptide and its therapeutic action. This  
XX sequence represents the human zinc finger protein 60.

XX Sequence 545 AA;

Query Match 41.4%; Score 53; DB 23; Length 545;

Best Local Similarity 50.0%; Pred. No. 89;

Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 5 PQQWTPEDQKAREAFR 22  
||:|:|:|:|:  
Db 186 PQSEKPYRDTEDREAFQ 203

RESULT 14  
AAU44633  
ID AAU44633 standard; Protein; 51 AA.

XX  
XX AAU44633;

XX 27-FEB-2002 (first entry)

XX Propionibacterium acnes immunogenic protein #5529.

XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
XX uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;

KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KW dermatological; osteopathic; neuroprotectant.

OS Propionibacterium acnes.

PN WO200181581-A2.

XX 01-NOV-2001.

XX 20-APR-2001; 2001WO-US12865.

XX 21-APR-2000; 2000US-199047P.

PR 02-JUN-2000; 2000US-208941P.

XX 07-JUL-2000; 2000US-216747P.

PA (CORI-) CORIXA CORP.

PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;

PI L'maisonneuve J, Zhang Y, Jen S, Carter D;

XX N-PSDB; AAS59523.

PT Propionibacterium acnes polypeptides and nucleic acids useful for  
 PT vaccinating against and diagnosing infections, especially useful for  
 PT treating acne vulgaris -

XX Example 1; SEQ ID No 5828; 1069pp; English.

CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
 CC polypeptides. The proteins and their associated DNA sequences are used in  
 CC the treatment, prevention and diagnosis of medical conditions caused by  
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
 CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.  
 CC P. acnes is also involved in infections of bone, joints and the central  
 CC nervous system, however it is particularly involved in the inflammatory  
 CC lesions associated with acne vulgaris. A method for detecting the  
 CC presence or absence of P. acnes in a patient comprises contacting a  
 CC sample with a binding agent that binds to the proteins of the invention  
 CC and determining the amount of bound protein in the sample. The  
 CC polypeptides may be used as antigens in the production of antibodies  
 CC specific for P. acnes proteins. These antibodies can be used to  
 CC downregulate expression and activity of P. acnes polypeptides and  
 CC therefore treat P. acnes infections. The antibodies may also be used as  
 CC diagnostic agents for determining P. acnes presence, for example, by  
 CC enzyme linked immunosorbent assay (ELISA).

CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 51 AA;

Query Match 40.2%; Score 51.5; DB 22; Length 51;

Best Local Similarity 35.1%; Pred. No. 12;

Matches 13; Conservative 4; Mismatches 3; Indels 17; Gaps 2;

QY 4 IPQP-----QW-----TPEQKAREAFRR 23

Db 12 IPAPQCTHROWPACQDDVYSRSTPDEQGRQSKR 48

RESULT 15

AAG27027  
 ID AAG27027 standard; Protein; 150 AA.

XX AAG27027;

DT 17-OCT-2000 (first entry)

DE Zea mays protein fragment SEQ ID NO: 31706.

KW Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;

KW termination sequence; corn.

OS Zea mays subsp. mays.

PN EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.

PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139750.

PR 18-JUN-1999; 99US-0139763.

PR 21-JUN-1999; 99US-0139817.

PR 21-JUN-1999; 99US-0139899.

PR 23-JUN-1999; 99US-0140353.

PR 23-JUN-1999; 99US-0140354.

PR 24-JUN-1999; 99US-0140695.

PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.

PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 08-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 22-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 39.8%; Score 51; DB 21; Length 150;  
Best Local Similarity 60.0%; Pred. No. 42;  
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 9 WTPEDQKAREAFRR 23  
      |||||: | | |  
Db 56 WTPEDQLLRRAITR 70

Search completed: January 2, 2003, 13:44:25  
Job time : 10.254 secs

This Page Blank (uspto)



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2003, 13:42:53 ; Search time 3.06048 Seconds  
(without alignments)  
221.118 Million cell updates/sec

Title: US-09-818-066-34\_COPY\_80\_102

Perfect score: 128

Sequence: 1 PQEIPQWTPPEQKAREAFRR 23

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents\_AA:\*

- 1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/1aa/PCTUS\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep.\*

Pred. No. Is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	128	100.0	48	2	US-08-683-262B-68
2	128	100.0	48	4	US-09-361-707-68
3	128	100.0	161	2	US-08-683-262B-34
4	128	100.0	161	4	US-09-361-707-34
5	88	68.8	16	2	US-08-683-262B-18
6	88	68.8	16	4	US-09-361-707-18
7	50	39.1	528	4	US-08-928-213B-8
8	48	37.5	172	4	US-09-285-601-2
9	48	37.5	751	4	US-09-402-929-2
10	48	37.5	752	4	US-09-402-929-6
11	48	37.5	755	3	US-09-071-101-2
12	48	37.5	755	3	US-09-369-618-2
13	48	37.5	755	3	US-09-369-617-2
14	47	36.7	450	2	US-08-665-037-2
15	47	36.7	450	2	US-08-666-067-2
16	47	36.7	1346	2	US-08-732-870-2
17	47	36.7	1346	2	US-08-635-121-2
18	47	36.7	1958	1	US-07-945-283-2
19	46	35.9	24	5	PCT-US93-05640-40
20	46	35.9	667	2	US-08-718-661-2
21	45.5	35.5	485	6	5320958-2
22	45	35.2	21	4	US-08-630-915A-148
23	45	35.2	50	4	US-09-156-316-4
24	45	35.2	156	4	US-08-928-941D-4
25	45	35.2	156	4	US-08-928-941D-36
26	45	35.2	156	4	US-09-280-590A-4
27	45	35.2	156	4	US-09-280-590A-46

28	45	35.2	317	2	US-08-864-799-4	Sequence 4, Appli
29	45	35.2	506	2	US-08-820-170A-19	Sequence 19, Appl
30	45	35.2	506	3	US-09-055-699-19	Sequence 19, Appl
31	45	35.2	506	4	US-09-273-565-19	Sequence 19, Appl
32	45	35.2	506	4	US-09-565-538-19	Sequence 19, Appl
33	45	35.2	506	4	US-09-661-468-19	Sequence 19, Appl
34	45	35.2	2647	2	US-08-583-562B-8	Sequence 8, Appli
35	45	35.2	2647	2	US-08-779-113-8	Sequence 8, Appli
36	44	34.4	50	4	US-09-156-316-5	Sequence 5, Appli
37	44	34.4	50	4	US-09-156-316-6	Sequence 5, Appli
38	44	34.4	464	4	US-08-580-031A-14	Sequence 14, Appl
39	44	34.4	566	4	US-09-431-470-2	Sequence 2, Appli
40	44	34.4	619	1	US-08-465-746-2	Sequence 2, Appli
41	44	34.4	619	1	US-08-214-164-2	Sequence 2, Appli
42	44	34.4	619	2	US-08-467-852A-3	Sequence 3, Appli
43	44	34.4	619	2	US-08-246-636-2	Sequence 2, Appli
44	44	34.4	619	2	US-08-247-491A-3	Sequence 3, Appli
45	44	34.4	619	2	US-08-319-795-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1  
US-08-683-262B-68  
; Sequence 68, Application US/08683262B  
; Patent No. 5929220  
; GENERAL INFORMATION:  
; APPLICANT: Shuping Tong et al.  
; TITLE OF INVENTION: HEPADNAVIRUS RECEPTOR  
; NUMBER OF SEQUENCES: 75  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/683.262B  
; FILING DATE: 18-JUL-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fraser, Janis K.  
; REGISTRATION NUMBER: 31,819  
; REFERENCE/DOCKET NUMBER: 00786/287002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 542-5070  
; TELEFAX: (617) 542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 68:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 48 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
US-08-683-262B-68

Query Match 100.0%; Score 128; DB 2; Length 48;  
Best Local Similarity 100.0%; Pred. No. 1.8e-11;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PQEIPQWTPPEQKAREAFRR 23  
Db 12 PQEIPQWTPPEQKAREAFRR 34

RESULT 2  
US-09-361-707-68

```
; Sequence 68, Application US/09361707
; Patent No. 6258937
; GENERAL INFORMATION:
; APPLICANT: Tong, Shuping
; Li, Jisu
; Wands, Jack R.
; TITLE OF INVENTION: HEPA DNA VIRUS RECEPTOR
; NUMBER OF SEQUENCES: 117
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/361,707
; FILING DATE: 27-Jul-1999
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 08/683,262
; FILING DATE: 18-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Creason, Gary L.
; REGISTRATION NUMBER: 34,310
; REFERENCE/DOCKET NUMBER: 00786/287003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 48 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-09-361-707-68

Query Match 100.0%; Score 128; DB 4; Length 48;
Best Local Similarity 100.0%; Pred. No. 1.8e-11;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 POEIPQQTPTPEEDQKAREAFRR 23
12 POEIPQQTPTPEEDQKAREAFRR 34
12 POEIPQQTPTPEEDQKAREAFRR 34

RESULT 3
US-08-683-262B-34
; Sequence 34, Application US/08683262B
; Patent No. 5929220
; GENERAL INFORMATION:
; APPLICANT: Shuping Tong et al.
; TITLE OF INVENTION: HEPADNAVIRUS RECEPTOR
; NUMBER OF SEQUENCES: 75
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/683,262B
```

```
; FILING DATE: 18-JUL-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 31,819
; REFERENCE/DOCKET NUMBER: 00786/287002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 161 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-683-262B-34

Query Match 100.0%; Score 128; DB 2; Length 161;
Best Local Similarity 100.0%; Pred. No. 7.3e-11;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 POEIPQQTPTPEEDQKAREAFRR 23
80 POEIPQQTPTPEEDQKAREAFRR 102
Db 80 POEIPQQTPTPEEDQKAREAFRR 102

RESULT 4
US-09-361-707-34
; Sequence 34, Application US/09361707
; Patent No. 6258937
; GENERAL INFORMATION:
; APPLICANT: Tong, Shuping
; Li, Jisu
; Wands, Jack R.
; TITLE OF INVENTION: HEPA DNA VIRUS RECEPTOR
; NUMBER OF SEQUENCES: 117
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/361,707
; FILING DATE: 27-Jul-1999
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 08/683,262
; FILING DATE: 18-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Creason, Gary L.
; REGISTRATION NUMBER: 34,310
; REFERENCE/DOCKET NUMBER: 00786/287003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 161 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; SEQUENCE DESCRIPTION: SEQ ID NO: 34:
US-09-361-707-34

Query Match 100.0%; Score 128; DB 4; Length 161;
```

Best Local Similarity 100.0%; Pred. No. 7.3e-11;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PQEIPQWTPEDQKAREAFRR 23  
Db 80 PQEIPQWTPEDQKAREAFRR 102

## RESULT 5

US-08-683-262B-18  
Sequence 18, Application US/08683262B  
Patent No. 5929220

## GENERAL INFORMATION:

APPLICANT: Shuping Tong et al.

TITLE OF INVENTION: HEPADNAVIRUS RECEPTOR

NUMBER OF SEQUENCES: 75

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson P.C.

STREET: 225 Franklin Street

CITY: Boston

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02110-2804

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

## CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/683,262B

FILING DATE: 18-JUL-1996

CLASSIFICATION: 435

## ATTORNEY/AGENT INFORMATION:

NAME: Fraser, Janis K.

REGISTRATION NUMBER: 31,819

REFERENCE/DOCKET NUMBER: 00786/287002

## TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 542-8906

TELEX: 200154

## INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHARACTERISTICS:

LENGTH: 16 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-683-262B-18

Query Match 68.8%; Score 88; DB 2; Length 16;

Best Local Similarity 100.0%; Pred. No. 2.1e-06;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 QWTPEEQKAREAFRR 23

Db 1 QWTPEEQKAREAFRR 16

## RESULT 6

US-09-361-707-18

Sequence 18, Application US/09361707

Patent No. 6258937

## GENERAL INFORMATION:

APPLICANT: Tong, Shuping

Li, Jisu

Wands, Jack R.

TITLE OF INVENTION: HEPA DNA VIRUS RECEPTOR

NUMBER OF SEQUENCES: 117

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson P.C.

STREET: 225 Franklin Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/361,707  
FILING DATE: 27-Jul-1999  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: 08/683,262  
FILING DATE: 18-JUL-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Creason, Gary L.  
REGISTRATION NUMBER: 34,310  
REFERENCE/DOCKET NUMBER: 00786/287003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 16 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 18:  
US-09-361-707-18

Query Match 68.8%; Score 88; DB 4; Length 16;

Best Local Similarity 100.0%; Pred. No. 2.1e-06;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 QWTPEEQKAREAFRR 23

Db 1 QWTPEEQKAREAFRR 16

## RESULT 7

US-08-928-213B-8

Sequence 8, Application US/08928213B

Patent No. 6238905

## GENERAL INFORMATION:

APPLICANT: McHenry, Charles S.

Seville, Mark

Cull, Millard G.

TITLE OF INVENTION: NOVEL THERMOPHILIC POLYMERASE III

HOLENZYME

NUMBER OF SEQUENCES: 195

CORRESPONDENCE ADDRESS:

ADDRESSEE: MEDLEN & CARROLL, LLP

STREET: 220 Montgomery Street, Suite 2200

CITY: San Francisco

STATE: CA

COUNTRY: USA

ZIP: 94104

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

## CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/928,213B

FILING DATE: 12-Sep-1997

CLASSIFICATION: <Unknown>

## ATTORNEY/AGENT INFORMATION:

NAME: MacKnight, Kamrin T.

REGISTRATION NUMBER: 38,230

REFERENCE/DOCKET NUMBER: ENZYCO-02550

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-705-8410

TELEFAX: 415-397-8338

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

TELEPHONE: (215) 568-5549  
TELEFAX: (215) 568-5549  
INFORMATION FOR SEQ ID NO:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 752 amino acids  
TYPE: amino acid  
STRANDEDNESS: single

```
; TOPOLOGY: linear
US-09-402-929-6

Query Match          37.5%: Score 48; DB 4; Length 752;
Best Local Similarity 42.9%: Pred. No. 67;
Matches 9; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

Qy 3 EIPQOWTPEDQKAREAFRR 23
| : ||| : || :
Db 136 EVKSSWTEEDRIIYEAHR 156

RESULT 11
US-09-071-101-2
; Sequence 2, Application US/09071101
; Patent No. 6013503
; GENERAL INFORMATION:
; APPLICANT: Lok, S1
; APPLICANT: Jaspers, Stephen R.
; TITLE OF INVENTION: HUMAN PROHORMONE CONVERTASE 4
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ZymoGenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071.101
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, Gary E.
; REGISTRATION NUMBER: 31,648
; REFERENCE/DOCKET NUMBER: 97-05
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-442-6673
; TELEFAX: 206-442-6678
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 755 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-09-071-101-2

Query Match          37.5%: Score 48; DB 3; Length 755;
Best Local Similarity 38.2%: Pred. No. 67;
Matches 13; Conservative 2; Mismatches 3; Indels 16; Gaps 2;

Qy 6 QPQ-----WTPEDQK-----AREAFRR 23
| | | | | | | | | | | | | | | |
Db 248 QPQHIIYSASWGPEDDGRVTDGPGILTREAFRR 281

RESULT 12
US-09-369-618-2
; Sequence 2, Application US/09369618
; Patent No. 6100041
; GENERAL INFORMATION:
; APPLICANT: Lok, S1
```

; CITY: Los Angeles  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 90071-2066  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; MEDIUM TYPE: storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: IBM P.C. DOS 5.0  
; SOFTWARE: Word Perfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/665,037  
; FILING DATE: June 13, 1996  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/005,167  
; FILING DATE: October 13, 1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warburg, Richard J.  
; REGISTRATION NUMBER: 32,327  
; REFERENCE/DOCKET NUMBER: 220/156  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (213) 489-1600  
; TELEFAX: (213) 955-0440  
; TELEEX: 67-3510  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 450 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-665-037-2

Query Match 36.7%; Score 47; DB 2; Length 450;  
Best Local Similarity 87.5%; Pred. No. 52;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 EIQPQWT 10  
I:|||||  
Db 360 ELPQPWT 367

RESULT 15  
US-08-666-067-2  
; Sequence 2, Application US/08666067  
; Patent No. 5922842  
; GENERAL INFORMATION:  
; APPLICANT: Seedorf, Klaus  
; APPLICANT: Ullrich, Axel  
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT  
; TITLE OF INVENTION: OF TKA-1 RELATED  
; TITLE OF INVENTION: DISORDERS  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lyon & Lyon  
; STREET: 633 West Fifth Street  
; STREET: Suite 4700  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 90071-2066  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; MEDIUM TYPE: storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: IBM P.C. DOS 5.0  
; SOFTWARE: Word Perfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/666,067  
; FILING DATE: June 13, 1996  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/005,421

; FILING DATE: October 13, 1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warburg, Richard J.  
; REGISTRATION NUMBER: 32,327  
; REFERENCE/DOCKET NUMBER: 220/157  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (213) 489-1600  
; TELEFAX: (213) 955-0440  
; TELEEX: 67-3510  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 450 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-666-067-2

Query Match 36.7%; Score 47; DB 2; Length 450;  
Best Local Similarity 87.5%; Pred. No. 52;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 EIQPQWT 10  
I:|||||  
Db 360 ELPQPWT 367

Search completed: January 2, 2003, 13:47:32  
Job time : 4.06048 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2003, 13:45:43 ; Search time 1.94758 Seconds  
(without alignments)  
223.808 Million cell updates/sec

Title: us-09-818-066-34\_COPY\_80\_102

Perfect score: 128

Sequence: 1 PQEIPQQTPEEDQKAREAFRR 23

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 117078 seqs, 18951520 residues

Total number of hits satisfying chosen parameters: 117078

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCCT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/2/pubpaa/PCCTUS\_PUBCOMB.pep:\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep:\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	128	100.0	48	10	US-09-818-066-68
2	128	100.0	161	10	US-09-818-066-34
3	88	68.8	16	10	US-09-818-066-18
4	53	41.4	485	10	US-09-864-761-37026
5	52	40.6	491	10	US-09-862-658-4
6	50	39.1	172	12	US-10-109-885-4
7	49	38.3	26	10	US-09-864-761-38935
8	49	38.3	216	10	US-09-884-441-484
9	49	38.3	469	10	US-09-828-303-22
10	48	37.5	755	10	US-09-888-615-82
11	47	36.7	203	10	US-09-925-299-872
12	47	36.7	807	9	US-10-138-221-7
13	47	36.7	1346	9	US-09-902-432-2
14	47	36.7	1596	9	US-09-902-432-4
15	46.5	36.3	364	9	US-09-978-295A-515
16	46.5	36.3	364	9	US-09-978-697-515
17	46.5	36.3	364	9	US-09-978-192A-515
18	46.5	36.3	364	9	US-09-999-832A-515
19	46.5	36.3	364	12	US-10-052-586-36

20	46.5	36.3	382	10	US-09-907-479-4
21	46	35.9	190	10	US-09-815-242-11013
22	46	35.9	249	9	US-09-738-626-6183
23	46	35.9	316	10	US-09-764-864-967
24	46	35.9	336	10	US-09-745-763-17
25	46	35.9	336	10	US-09-799-777-24
26	46	35.9	26926	9	US-09-759-5088-2
27	45.5	35.5	342	9	US-09-738-626-5165
28	45.5	35.5	485	8	US-08-808-031A-2
29	45	35.2	21	10	US-09-879-957-148
30	45	35.2	50	10	US-09-757-049A-4
31	45	35.2	193	10	US-09-920-552-102
32	45	35.2	305	9	US-10-008-118A-8
33	45	35.2	305	10	US-09-443-704-8
34	45	35.2	506	10	US-09-976-165-19
35	45	35.2	559	10	US-09-877-633-12
36	45	35.2	1814	10	US-09-920-552-103
37	44	34.4	50	10	US-09-757-049A-5
38	44	34.4	50	10	US-09-757-049A-6
39	44	34.4	464	12	US-10-095-492-14
40	43	33.6	51	10	US-09-912-962-16
41	43	33.6	92	10	US-09-925-301-1590
42	43	33.6	116	9	US-10-008-118A-40
43	43	33.6	116	10	US-09-443-704-40
44	43	33.6	189	10	US-09-785-738A-8
45	43	33.6	417	10	US-09-963-285-4

## ALIGNMENTS

RESULT 1  
US-09-818-066-68  
Sequence 68, Application US/09818066  
Patent No. US20020032307A1  
GENERAL INFORMATION:  
APPLICANT: Shuping Tong et al.  
TITLE OF INVENTION: HEPADNAVIRUS RECEPTOR  
NUMBER OF SEQUENCES: 75  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/818.066  
FILING DATE: 27-Mar-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/683,262  
FILING DATE: 18-JUL-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Fraser, Janis K.  
REGISTRATION NUMBER: 31,819  
REFERENCE/DOCKET NUMBER: 00786/287002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 68:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 48 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 68:  
US-09-818-066-68

Sequence 4, Appli  
Sequence 11013, A  
Sequence 6183, Ap  
Sequence 967, App  
Sequence 17, Appli  
Sequence 24, Appli  
Sequence 2, Appli  
Sequence 5165, Ap  
Sequence 2, Appli  
Sequence 148, App  
Sequence 4, Appli  
Sequence 102, App  
Sequence 8, Appli  
Sequence 19, Appli  
Sequence 12, Appli  
Sequence 103, App  
Sequence 5, Appli  
Sequence 6, Appli  
Sequence 14, Appli  
Sequence 16, Appli  
Sequence 1590, Ap  
Sequence 40, Appli  
Sequence 40, Appli  
Sequence 8, Appli  
Sequence 4, Appli

Query Match 100.0%; Score 128; DB 10; Length 48;  
Best Local Similarity 100.0%; Pred. No. 3.2e-11;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 POEIPQPTPEEDQKAREAFRR 23  
DB 12 POEIPQPTPEEDQKAREAFRR 34  
|||||

RESULT 2  
US-09-818-066-34  
; Sequence 34, Application US/09818066  
; Patent No. US20020032307A1

; GENERAL INFORMATION:  
; APPLICANT: Shuping Tong et al.  
; TITLE OF INVENTION: HEPADNAVIRUS RECEPTOR  
; NUMBER OF SEQUENCES: 75  
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02110-2804

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/818,066  
; FILING DATE: 27-Mar-2001  
; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/683,262  
; FILING DATE: 18-JUL-1996

; ATTORNEY/AGENT INFORMATION:  
; NAME: Fraser, Janis K.  
; REGISTRATION NUMBER: 31,819  
; REFERENCE/DOCKET NUMBER: 00786/287002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 542-5070  
; TELEFAX: (617) 542-8906  
; TELEX: 200154

; INFORMATION FOR SEQ ID NO: 34:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 161 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 34:  
US-09-818-066-34

Query Match 100.0%; Score 128; DB 10; Length 161;  
Best Local Similarity 100.0%; Pred. No. 1.3e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 POEIPQPTPEEDQKAREAFRR 23  
DB 80 POEIPQPTPEEDQKAREAFRR 102  
|||||

RESULT 3  
US-09-818-066-18  
; Sequence 18, Application US/09818066  
; Patent No. US20020032307A1

; GENERAL INFORMATION:  
; APPLICANT: Shuping Tong et al.  
; TITLE OF INVENTION: HEPADNAVIRUS RECEPTOR  
; NUMBER OF SEQUENCES: 75  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.

; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/818,066  
; FILING DATE: 27-Mar-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/683,262  
; FILING DATE: 18-JUL-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fraser, Janis K.  
; REGISTRATION NUMBER: 31,819  
; REFERENCE/DOCKET NUMBER: 00786/287002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 542-5070  
; TELEFAX: (617) 542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 16 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 18:  
US-09-818-066-18

Query Match 68.8%; Score 88; DB 10; Length 16;  
Best Local Similarity 100.0%; Pred. No. 2.4e-06;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 QWTPEEDQKAREAFRR 23  
DB 1 QWTPEEDQKAREAFRR 16  
|||||

RESULT 4  
US-09-864-761-37026  
; Sequence 37026, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Acomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30



;; PRIOR APPLICATION NUMBER: PCT/US01/00669  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00665  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00668  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00663  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00662  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00661  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00670  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: US 60/234,687  
;; PRIOR FILING DATE: 2000-09-21  
;; PRIOR APPLICATION NUMBER: US 09/608,408  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: US 09/774,203  
;; PRIOR FILING DATE: 2001-01-29  
;; NUMBER OF SEQ ID NOS: 49117  
;; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1  
;; SEQ ID NO 37026  
;; LENGTH: 485  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; OTHER INFORMATION: MAP TO AC003002.1  
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.5  
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2  
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.5  
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.5  
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.2  
;; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.5  
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.5  
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.2  
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.8  
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.3  
;; OTHER INFORMATION: SWISSPROT HIT: P52740, EVALUATE 1.00e-127  
;; OTHER INFORMATION: EST\_HUMAN HIT: AV683451.1, EVALUATE 1.00e-120  
US-09-864-761-37026

Query Match 41.48; Score 53; DB 10; Length 485;  
Best Local Similarity 50.08; Pred. No. 6.1;  
Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 5 PQQWTPEDQKAREAF 22  
||:|:|:|:|:|:|:  
b 126 PQSEKKPYRDTEDREAFQ 143

RESULT 5  
US-09-862-658-4  
;; Sequence 4, Application US/09862658  
;; Patent No. US20020137101A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Meyers, Rachel  
;; TITLE OF INVENTION: 46638, A NOVEL HUMAN LIPOXYGENASE FAMILY  
;; TITLE OF INVENTION: MEMBER AND USES THEREOF  
;; FILE REFERENCE: 10448-053001  
;; CURRENT APPLICATION NUMBER: US/09/862,658  
;; CURRENT FILING DATE: 2001-05-22  
;; PRIOR APPLICATION NUMBER: US 60/205,675  
;; PRIOR FILING DATE: 2000-05-19  
;; NUMBER OF SEQ ID NOS: 6  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 4  
;; LENGTH: 491  
;; TYPE: PRT  
;; ORGANISM: Artificial Sequence  
;; FEATURE:  
;; OTHER INFORMATION: consensus sequence  
US-09-862-658-4

Query Match 40.6%; Score 52; DB 10; Length 491;  
Best Local Similarity 55.6%; Pred. No. 8.5;  
Matches 10; Conservative 4; Mismatches 2; Indels 2; Gaps 1;

Qy 6 PQQWTPEDQKAREAFRR 23  
||:|:|:|:|:|:|:  
Db 438 EPEWT--SDKKALEAFKR 453

;; RESULT 6  
;; US-10-109-885-4  
;; Sequence 4, Application US/10109885  
;; Patent No. US20020119129A1  
;; GENERAL INFORMATION:  
;; APPLICANT: REVEL, Michel  
;; APPLICANT: CHEBATH, Judith  
;; APPLICANT: ABRAMOVITCH, Carolina  
;; TITLE OF INVENTION: NOVEL IEN RECEPTOR I BINDING PROTEIN, DNA ENCODING THEM, AND M  
;; TITLE OF INVENTION: MODULATING CELLULAR RESPONSE TO INTERFERON  
;; FILE REFERENCE: REVEL-14A  
;; CURRENT APPLICATION NUMBER: US/10/109,885  
;; CURRENT FILING DATE: 2002-04-01  
;; PRIOR APPLICATION NUMBER: US/09/341,640  
;; PRIOR FILING DATE: 1999-10-18  
;; PRIOR APPLICATION NUMBER: PCT/US98/00671  
;; PRIOR FILING DATE: 1998-01-15  
;; PRIOR APPLICATION NUMBER: US 60/035,636  
;; PRIOR FILING DATE: 1997-01-15  
;; NUMBER OF SEQ ID NOS: 13  
;; SOFTWARE: PatentIn version 3.1  
;; SEQ ID NO 4  
;; LENGTH: 172  
;; TYPE: PRT  
;; ORGANISM: Artificial Sequence  
;; FEATURE:  
;; OTHER INFORMATION: synthetic  
US-10-109-885-4

Query Match 39.18; Score 50; DB 12; Length 172;  
Best Local Similarity 52.9%; Pred. No. 4.9;  
Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 5 PQQWTPEDQKAREAF 21  
||:|:|:|:|:|:|:  
Db 21 KPPELTERQKQREAF 37

;; RESULT 7  
;; US-09-864-761-38935  
;; Sequence 38935, Application US/09864761  
;; Patent No. US20020048763A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Penn, Sharron G.  
;; APPLICANT: Rank, David R.  
;; APPLICANT: Hanzel, David K.  
;; APPLICANT: Chen, Wensheng  
;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO  
;; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
;; FILE REFERENCE: Acomica-X-1  
;; CURRENT APPLICATION NUMBER: US/09/864,761  
;; CURRENT FILING DATE: 2001-05-23  
;; PRIOR APPLICATION NUMBER: US 60/180,312  
;; PRIOR FILING DATE: 2000-02-04  
;; PRIOR APPLICATION NUMBER: US 60/207,456  
;; PRIOR FILING DATE: 2000-05-26  
;; PRIOR APPLICATION NUMBER: US 09/632,366  
;; PRIOR FILING DATE: 2000-08-03  
;; PRIOR APPLICATION NUMBER: GB 24263.6  
;; PRIOR FILING DATE: 2000-10-04  
;; PRIOR APPLICATION NUMBER: US 60/236,359  
;; PRIOR FILING DATE: 2000-09-27  
;; PRIOR APPLICATION NUMBER: PCT/US01/00666

;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00667  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00664  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00669  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00665  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00668  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00663  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00662  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00661  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00670  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: US 60/234,687  
;; PRIOR FILING DATE: 2000-09-21  
;; PRIOR APPLICATION NUMBER: US 09/608,408  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: US 09/774,203  
;; PRIOR FILING DATE: 2001-01-29  
;; NUMBER OF SEQ ID NOS: 49117  
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
;; SEQ ID NO 38935  
;; LENGTH: 26  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; OTHER INFORMATION: MAP TO AC005822.1  
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.7  
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.6  
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5  
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.98  
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.3  
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.3  
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4  
;; OTHER INFORMATION: EST\_HUMAN HIT: AW404800.1, EVALUATE 1.00e-07  
US-09-864-761-38935

Query Match 38.3%; Score 49; DB 10; Length 26;  
Best Local Similarity 45.5%; Pred. No. 0.79;  
Matches 10; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

OY 1 POEIPQOWTPEEDQKAREAFR 22  
||| | ||||| : : ||| :  
1 PONSTLSQDTPEEDPRGRHAFQ 22

RESULT 8  
US-09-884-441-484  
; Sequence 484, Application US/09884441  
; Patent No. US20020119158A1  
; GENERAL INFORMATION:  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Carter, Darrick  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
; TITLE OF INVENTION: DIAGNOSIS OF OVARIAN CANCER  
; FILE REFERENCE: 210121.462C7  
; CURRENT APPLICATION NUMBER: US/09/884,441  
; CURRENT FILING DATE: 2001-06-18  
; NUMBER OF SEQ ID NOS: 489  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 484  
; LENGTH: 216  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-884-441-484

Query Match 38.3%; Score 49; DB 10; Length 216;

Best Local Similarity 72.7%; Pred. No. 8.6;  
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
OY 1 POEIPQOWTPT 11  
||| | : |||||  
DB 36 POWISEPQWTP 46

RESULT 9  
US-09-828-303-22  
; Sequence 22, Application US/09828303  
; Patent No. US20020102695A1  
; GENERAL INFORMATION:  
; APPLICANT: COSTA E SILVA, OSWALDO DA  
; APPLICANT: BOHNERT, HANS J.  
; APPLICANT: VAN THIELEN, NOCHA  
; APPLICANT: CHEN, ROUYING  
; TITLE OF INVENTION: TRANSCRIPTION FACTOR STRESS-RELATED PROTEINS AND  
; TITLE OF INVENTION: METHODS OF USE IN PLANTS  
; FILE REFERENCE: 16313-0030  
; CURRENT APPLICATION NUMBER: US/09/828,303  
; CURRENT FILING DATE: 2001-08-20  
; PRIOR APPLICATION NUMBER: 60/196,001  
; PRIOR FILING DATE: 2000-04-07  
; NUMBER OF SEQ ID NOS: 79  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 22  
; LENGTH: 469  
; TYPE: PRT  
; ORGANISM: Physcomitrella patens  
US-09-828-303-22

Query Match 38.3%; Score 49; DB 10; Length 469;  
Best Local Similarity 44.4%; Pred. No. 20;  
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

OY 2 QEIPQOWTPEEDQKARE 19  
:: | : | ||| |  
DB 22 KQCPRGHWRPAEDDKLRE 39

RESULT 10  
US-09-888-615-82  
; Sequence 82, Application US/09888615  
; Patent No. US20020064856A1  
; GENERAL INFORMATION:  
; APPLICANT: PLOWMAN, GREGORY  
; APPLICANT: WHYTE, DAVID  
; APPLICANT: CAENEPEEL, SEAN  
; APPLICANT: CHARYDCZAK, GLEN  
; APPLICANT: MANNING, GERARD  
; APPLICANT: SUDARSANAM, SUCHA  
; TITLE OF INVENTION: NOVEL PROTEASES  
; FILE REFERENCE: 038602/1214  
; CURRENT APPLICATION NUMBER: US/09/888,615  
; CURRENT FILING DATE: 2001-06-26  
; PRIOR APPLICATION NUMBER: 60/214,047  
; PRIOR FILING DATE: 2000-06-26  
; NUMBER OF SEQ ID NOS: 150  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 82  
; LENGTH: 755  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-888-615-82

Query Match 37.5%; Score 48; DB 10; Length 755;  
Best Local Similarity 38.2%; Pred. No. 48;  
Matches 13; Conservative 2; Mismatches 3; Indels 16; Gaps 2;

OY 6 QPQHIIYSASWGPDDGRTVDGPGLITREAFRR 23  
||| | ||| :  
DB 248 QPQHIIYSASWGPDDGRTVDGPGLITREAFRR 281

```
RESULT 11
US-09-925-299-872
; Sequence 872, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 872
; LENGTH: 203
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-299-872

Query Match          36.7%; Score 47; DB 10; Length 203;
Best Local Similarity 44.4%; Pred. No. 15;
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 2 QEIPQQTPEEDQKARE 19
   ||||| ||||| |
Db 119 EEPKPEEKPEEKLEE 136

RESULT 12
US-10-138-221-7
; Sequence 7, Application US/10138221
; Publication No. US2002019216A1
; GENERAL INFORMATION:
; APPLICANT: Amy F. MacRae
; TITLE OF INVENTION: USE OF TRANSPOSABLE ELEMENTS FOR ALTERING GENE EXPRESSION
; FILE REFERENCE: 511/877
; CURRENT APPLICATION NUMBER: US/10/138,221
; CURRENT FILING DATE: 2002-08-01
; PRIOR APPLICATION NUMBER: US 60/287,882
; PRIOR FILING DATE: 2001-05-01
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 7
; LENGTH: 807
; TYPE: PRT
; ORGANISM: Zea mays
US-10-138-221-7

Query Match          36.7%; Score 47; DB 9; Length 807;
Best Local Similarity 52.9%; Pred. No. 70;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 PQEIPQQTPEEDQKA 17
   || ||||| |||||
Db 115 PQEPQPPQPEPEEEA 131

RESULT 13
US-09-902-432-2
; Sequence 2, Application US/09902432
; Patent No. US20020160002A1
; GENERAL INFORMATION:
; APPLICANT: Irwin H. Gelman
; APPLICANT: Susan G. Jaken
; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE
; FILE REFERENCE: A30558-A-FWC-A 070156.0597
; CURRENT APPLICATION NUMBER: US/09/902,432
; CURRENT FILING DATE: 2002-04-08

; PRIOR APPLICATION NUMBER: 08/978,277
; PRIOR FILING DATE: 1997-11-25
; PRIOR APPLICATION NUMBER: 08/665,401
; PRIOR FILING DATE: 1996-06-18
; PRIOR APPLICATION NUMBER: 08/635,121
; PRIOR FILING DATE: 1996-04-19
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1596
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-09-902-432-2

Query Match          36.7%; Score 47; DB 9; Length 1596;
Best Local Similarity 52.6%; Pred. No. 1.5e+02;
Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 1 PQEIPQQTPEEDQKARE 19
   ||||| ||||| |||||
Db 436 PQEVPQEAPEAEELMKSRE 454

RESULT 14
US-09-902-432-4
; Sequence 4, Application US/09902432
; Patent No. US20020160002A1
; GENERAL INFORMATION:
; APPLICANT: Irwin H. Gelman
; APPLICANT: Susan G. Jaken
; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE
; FILE REFERENCE: A30558-A-FWC-A 070156.0597
; CURRENT APPLICATION NUMBER: US/09/902,432
; CURRENT FILING DATE: 2002-04-08
; PRIOR APPLICATION NUMBER: 08/978,277
; PRIOR FILING DATE: 1997-11-25
; PRIOR APPLICATION NUMBER: 08/665,401
; PRIOR FILING DATE: 1996-06-18
; PRIOR APPLICATION NUMBER: 08/635,121
; PRIOR FILING DATE: 1996-04-19
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1596
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-09-902-432-4

Query Match          36.7%; Score 47; DB 9; Length 1346;
Best Local Similarity 52.6%; Pred. No. 1.2e+02;
Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 1 PQEIPQQTPEEDQKARE 19
   ||||| ||||| |||||
Db 50 PQEVPQEAPEAEELMKSRE 58

RESULT 15
US-09-978-295A-515
; Sequence 515, Application US/09978295A
; Patent No. US20020156006A1
; GENERAL INFORMATION:
; APPLICANT: ASHkenazi, Avi
; APPLICANT: Baker Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan
; APPLICANT: Ferrara, Napoleon
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
```

APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Grimaldi, J. Christopher  
APPLICANT: Gurney, Austin L.  
APPLICANT: Hillan, Kenneth J.  
APPLICANT: Kljavin, Ivar J.  
APPLICANT: Kuo, Sophia S.  
APPLICANT: Napier, Mary A.  
APPLICANT: Pan, James  
APPLICANT: Paoni, Nicholas F.  
APPLICANT: Roy, Margaret Ann  
APPLICANT: Shelton, David L.  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Williams, P. Mickey  
APPLICANT: Wood, William I.  
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
TITLE OF INVENTION: Acids Encoding the Same  
FILE REFERENCE: P2630PIC11  
CURRENT APPLICATION NUMBER: US/09/978,295A  
CURRENT FILING DATE: 2001-10-15  
PRIOR APPLICATION NUMBER: 09/918585  
PRIOR FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: 60/062250  
PRIOR FILING DATE: 1997-10-17  
PRIOR APPLICATION NUMBER: 60/064249  
PRIOR FILING DATE: 1997-11-03  
PRIOR APPLICATION NUMBER: 60/065311  
PRIOR FILING DATE: 1997-11-13  
PRIOR APPLICATION NUMBER: 60/066364  
PRIOR FILING DATE: 1997-11-21  
PRIOR APPLICATION NUMBER: 60/077450  
PRIOR FILING DATE: 1998-03-10  
PRIOR APPLICATION NUMBER: 60/077632  
PRIOR FILING DATE: 1998-03-11  
PRIOR APPLICATION NUMBER: 60/077641  
PRIOR FILING DATE: 1998-03-11  
PRIOR APPLICATION NUMBER: 60/077649  
PRIOR FILING DATE: 1998-03-11  
PRIOR APPLICATION NUMBER: 60/077791  
PRIOR FILING DATE: 1998-03-12  
PRIOR APPLICATION NUMBER: 60/078004  
PRIOR FILING DATE: 1998-03-13  
PRIOR APPLICATION NUMBER: 60/078886  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/078936  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/078910  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/078939  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/079294  
PRIOR FILING DATE: 1998-03-25  
PRIOR APPLICATION NUMBER: 60/079656  
PRIOR FILING DATE: 1998-03-26  
PRIOR APPLICATION NUMBER: 60/079664  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079689  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079663  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079728  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079786  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079920  
PRIOR FILING DATE: 1998-03-30  
PRIOR APPLICATION NUMBER: 60/079923  
PRIOR FILING DATE: 1998-03-30  
PRIOR APPLICATION NUMBER: 60/080105  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080107  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080165  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080194  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080327  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/080328  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/080333  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/080334  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/081070  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081049  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081071  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081195  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081203  
PRIOR FILING DATE: 1998-04-09  
PRIOR APPLICATION NUMBER: 60/081229  
PRIOR FILING DATE: 1998-04-09  
PRIOR APPLICATION NUMBER: 60/081955  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081817  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081819  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081952  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081838  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/082568  
PRIOR FILING DATE: 1998-04-21  
PRIOR APPLICATION NUMBER: 60/082569  
PRIOR FILING DATE: 1998-04-21  
PRIOR APPLICATION NUMBER: 60/082704  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/082804  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/082700  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/082797  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/082796  
PRIOR FILING DATE: 1998-04-23  
PRIOR APPLICATION NUMBER: 60/083336  
PRIOR FILING DATE: 1998-04-27  
PRIOR APPLICATION NUMBER: 60/083322  
PRIOR FILING DATE: 1998-04-28  
PRIOR APPLICATION NUMBER: 60/083392  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083495  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083496  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083499  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083545  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083554  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083558  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083559  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083500  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083742  
PRIOR FILING DATE: 1998-04-30  
PRIOR APPLICATION NUMBER: 60/084366  
PRIOR FILING DATE: 1998-03-31



This Page Blank (uspto)

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2003, 13:41:48 ; Search time 3.43145 Seconds  
(without alignments)  
644.360 Million cell updates/sec

Title: US-09-818-066-34\_COPY\_80\_102

Perfect score: 128

Sequence: 1 PQEIPQQTPEEDQKAREAFRR 23

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_73.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	128	100.0	158	2 JC1095	Pre-S protein - du
2	128	100.0	364	1 SAVLD	large surface anti
3	128	100.0	365	1 SAVLWE	large surface anti
4	121	94.5	366	1 SAVLWD	large surface anti
5	118	92.2	366	1 SAVLWD	large surface anti
6	62.5	48.8	335	1 SAVLHH	large surface anti
7	56	43.8	311	2 AC0563	conserved hypotet
8	54	42.2	390	2 S67439	hypothetical prote
9	54	42.2	1112	2 D75056	cell division cont
10	52	40.6	744	2 F97364	malate synthase G
11	52	40.6	744	2 AH2582	malate synthase G
12	51.5	40.2	357	2 AH0461	sn-glycerol-3-phos
13	51.5	40.2	776	2 E85384	probable myb-prote
14	51	39.8	1355	2 T22552	hypothetical prote
15	50.5	39.5	733	1 S33643	transforming prote
16	50	39.1	172	2 A49552	caltractin - human
17	50	39.1	624	1 I51581	transforming prote
18	50	39.1	986	2 T10754	cis-Golgi matrix p
19	50	39.1	1078	2 S77162	DNA topoisomerase
20	50	39.1	1297	2 T52065	probable myb-relat
21	49	38.3	115	2 E84512	probable MYB famil
22	49	38.3	117	2 F82308	hypothetical prote
23	49	38.3	148	2 I39576	ncx protein - Alc
24	49	38.3	324	2 B85064	MYB-like protein l
25	49	38.3	326	2 T49966	myb-related protei
26	49	38.3	352	2 T51659	myb-related transc
27	49	38.3	365	2 D86470	F21H2.9 protein -
28	49	38.3	405	2 JQ2147	OHPI protein - mai
29	49	38.3	405	2 JC5175	seed storage prote

30 49 38.3 421 1 S26605 myb-related protei  
31 49 38.3 529 2 T48253 myb-like protein -  
32 49 38.3 1193 2 T32016 hypothetical prote  
33 49 38.3 2225 1 A23443 pyrimidine synthe  
34 48 37.5 172 2 I38424 centr in - human  
35 48 37.5 172 2 S38531 caltractin - mouse  
36 48 37.5 301 2 C82791 conserved hypotet  
37 48 37.5 410 2 JQ2148 OHP2 protein - mai  
38 48 37.5 459 2 S51302 SIN3 protein-bindi  
39 48 37.5 613 2 A56031 potassium channel  
40 48 37.5 637 2 A45777 cell division cont  
41 48 37.5 751 1 I49497 transforming prote  
42 48 37.5 752 1 S03423 transforming prote  
43 48 37.5 802 2 A87754 protein C43E11.11  
44 48 37.5 865 1 S23454 lipoxigenase (EC 1  
45 48 37.5 876 2 T05943 probable lipoxigen

#### ALIGNMENTS

##### RESULT 1

JC1095

Pre-S protein - duck hepatitis virus

C;Species: duck hepatitis virus, DHBV

C;Date: 27-Aug-1995 #sequence\_revision 27-Oct-1995 #text\_change 09-May-1997

C;Accession: JC1095

R;Ma, Z.M.; Li, B.L.; Xiong, S.D.; Wen, Y.M.

Chinese J. Virol. 10, 1-7, 1994

A;Title: High expression of duck hepatitis B virus pre-S antigen in Escherichia coli

A;Reference number: JC1095

A;Accession: JC1095

A;Molecule type: DNA

A;Residues: 1-158 <MAX>

C;Genetics:

A;Gene: pre-S

C;Superfamily: hepatitis B virus surface antigen

C;Keywords: surface antigen

Query Match 100.0%; Score 128; DB 2; Length 158;

Best Local Similarity 100.0%; Pred. No. 8.2e-10;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PQEIPQQTPEEDQKAREAFRR 23

|||||

Db 109 PQEIPQQTPEEDQKAREAFRR 131

##### RESULT 2

SAVLD

large surface antigen - duck hepatitis virus

N;Contains: major surface antigen; middle surface antigen

C;Species: duck hepatitis virus, DHBV

C;Date: 20-Sep-1984 #sequence\_revision 20-Sep-1984 #text\_change 13-Mar-1997

C;Accession: A03710; S12845

R;Mandart, E.; Kay, A.; Galibert, F.

J. Virol. 49, 782-792, 1984

A;Title: Nucleotide sequence of a cloned duck hepatitis B virus genome: comparison wi

A;Reference number: A92997; MUID:84138772; PMID:6699938

A;Accession: A03710

A;Molecule type: DNA

A;Residues: 1-364 <MAN>

A;Cross-references: GB:K01834

R;Mates, F.; Tong, S.; Teubner, K.; Blum, H.E.

Nucleic Acids Res. 18, 6140, 1990

A;Title: Complete nucleotide sequence of a German duck hepatitis B virus.

A;Reference number: S12843; MUID:91045092; PMID:2235507

A;Accession: S12845

A;Status: translation not shown

A;Molecule type: DNA

A;Residues: 36-364 <MAT>

A;Cross-references: EMBL:X12798

C;Genetics:

A:Gene: pre-S1/pre-S2/S  
C:Superfamily: hepatitis B virus surface antigen  
C:Keywords: glycoprotein; surface antigen  
F:89-364/Product: middle surface antigen (gene pre-S2/S) #status predicted <DSA>  
F:198-364/Product: major surface antigen (gene S) #status predicted <MSA>  
F:32,170,296/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 128; DB 1; Length 364;  
Best Local Similarity 100.0%; Pred. No. 2e-09;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 POEIPQOWTPEEDQKAREAFRR 23  
DB 116 POEIPQOWTPEEDQKAREAFRR 138  
|||||

RESULT 3  
SAVLWD  
large surface antigen - duck hepatitis virus (strain China)  
N:Contains: major surface antigen; middle surface antigen  
C:Species: duck hepatitis virus, DHBV  
A:Title: Complete nucleotide sequence of a Chinese duck hepatitis B virus.  
Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 16-Jul-1999  
C:Accession: S12842  
R:Tong, S.; Mattes, F.; Teubner, K.; Blum, H.E.  
Nucleic Acids Res. 18, 6139, 1990  
A:Title: Complete nucleotide sequence of a Chinese duck hepatitis B virus.  
A:Reference number: S12840; MUID:91045091; PMID:2235506  
A:Accession: S12842  
A:Status: translation not shown  
A:Molecule type: DNA  
A:Residues: 1-365 <TON>  
A:Cross-references: GB:M21953; NID:g325435; PIDN:AAA45746.1; PID:g325438  
C:Genetics:

A:Gene: pre-S1/pre-S2/S  
C:Superfamily: hepatitis B virus surface antigen  
C:Keywords: glycoprotein; surface antigen  
F:89-365/Product: middle surface antigen (gene pre-S2/S) #status predicted <DSA>  
F:199-365/Product: major surface antigen (gene S) #status predicted <MSA>  
F:297/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 128; DB 1; Length 365;  
Best Local Similarity 100.0%; Pred. No. 2e-09;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 POEIPQOWTPEEDQKAREAFRR 23  
DB 116 POEIPQOWTPEEDQKAREAFRR 138  
|||||

RESULT 4  
SAVLWD  
large surface antigen - duck hepatitis virus (strain S31)  
N:Contains: major surface antigen; middle surface antigen  
C:Species: duck hepatitis virus, DHBV  
A:Note: host (Shanghai white duck)  
C:Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 04-Oct-1996  
C:Accession: D33746  
R:Uchida, M.; Esumi, M.; Shikata, T.  
Virology 173, 600-606, 1989  
A:Title: Molecular cloning and sequence analysis of duck hepatitis B virus genomes of a  
A:Reference number: A33746; MUID:90085807; PMID:2596031  
A:Accession: D33746  
A:Molecule type: DNA  
A:Residues: 1-366 <UCH>  
A:Cross-references: GB:M32990  
C:Genetics:

A:Gene: pre-S1/pre-S2/S  
C:Superfamily: hepatitis B virus surface antigen  
C:Keywords: glycoprotein; surface antigen  
F:89-366/Product: middle surface antigen (gene pre-S2/S) #status predicted <DSA>  
F:200-366/Product: major surface antigen (gene S) #status predicted <MSA>  
F:170,298/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 94.5%; Score 121; DB 1; Length 366;  
Best Local Similarity 95.7%; Pred. No. 1.7e-08;  
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 POEIPQOWTPEEDQKAREAFRR 23  
DB 116 POEIPQOWTPEEDQKAREAFRR 138  
|||||

RESULT 5  
SAVLBD  
large surface antigen - duck hepatitis virus (strain S5)  
N:Contains: major surface antigen; middle surface antigen  
C:Species: duck hepatitis virus, DHBV  
A:Note: host (Shanghai brown duck)  
C:Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 04-Oct-1996  
C:Accession: C33746  
R:Uchida, M.; Esumi, M.; Shikata, T.  
Virology 173, 600-606, 1989  
A:Title: Molecular cloning and sequence analysis of duck hepatitis B virus genomes of  
A:Reference number: A33746; MUID:90085807; PMID:2596031  
A:Accession: C33746  
A:Molecule type: DNA  
A:Residues: 1-366 <UCH>  
A:Cross-references: GB:M32990  
C:Genetics:

A:Gene: pre-S1/pre-S2/S  
C:Superfamily: hepatitis B virus surface antigen  
C:Keywords: glycoprotein; surface antigen  
F:89-366/Product: middle surface antigen (gene pre-S2/S) #status predicted <DSA>  
F:200-366/Product: major surface antigen (gene S) #status predicted <MSA>  
F:170,298/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 92.2%; Score 118; DB 1; Length 366;  
Best Local Similarity 91.3%; Pred. No. 4.1e-08;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 POEIPQOWTPEEDQKAREAFRR 23  
DB 116 PHETPQOWTPEEDQKAREAFRR 138  
|||||

RESULT 6  
SAVLHH  
large surface antigen - heron hepatitis virus  
N:Contains: major surface antigen; middle surface antigen  
C:Species: heron hepatitis virus, HHV  
A:Note: host Ardea cinerea (gray heron)  
C:Date: 30-Sep-1989 #sequence\_revision 30-Sep-1989 #text\_change 16-Jul-1999  
C:Accession: B30082  
R:Spengel, R.; Kaleta, E.F.; Will, H.  
J. Virol. 62, 3832-3839, 1988  
A:Title: Isolation and characterization of a hepatitis B virus endemic in herons.  
A:Reference number: A93037; MUID:88333160; PMID:3418788  
A:Accession: B30082  
A:Molecule type: DNA  
A:Residues: 1-335 <SPR>  
A:Cross-references: GB:M22056; NID:g325452; PIDN:AAA45739.1; PID:g325455  
C:Genetics:

A:Gene: pre-S1/pre-S2/S  
C:Superfamily: hepatitis B virus surface antigen  
C:Keywords: glycoprotein; surface antigen  
F:130-335/Product: middle surface antigen (gene pre-S2/S) #status predicted <DSA>  
F:167-335/Product: major surface antigen (gene S) #status predicted <MSA>  
F:265/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 48.8%; Score 62.5; DB 1; Length 335;  
Best Local Similarity 50.0%; Pred. No. 0.75;  
Matches 12; Conservative 5; Mismatches 6; Indels 1; Gaps 1;

QY 1 POEIPQOWTPEEDQKAREAFRR 23  
DB 81 PRPTPPPTWTTEEDKAKAEFFKQ 104  
|||||



A;Accession: AH2582  
A;Status: preliminary

A:Molecule type: DNA  
A:Residues: 1-744 <KUR>  
A:Cross-references: GB:AE008688; PIDN:AAL41078.1; PID:gl7738367; GSPDB:GN00186  
A:Experimental source: strain C58 (Dupont)  
C:Genetics:  
A:Gene: glcB  
A:Map position: circular chromosome

Query Match 40.6%; Score 52; DB 2; Length 744;  
Best Local Similarity 66.7%; Pred. NO. 42;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 5 PQQWTPPEEDQK 16  
DB 602 PRPWTPEEIQR 613

RESULT 12  
AH0461

A:Title: glycerol-3-phosphate transport, ATP-binding protein ugpC [imported] - Yersinia pestis  
A:Species: Yersinia pestis  
A:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 19-Jul-2002  
C:Accession: AH0461  
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; Hill, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, Nature 413, 523-527, 2001  
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A:Reference number: AB0001; MUID:21470413; PMID:11586360  
A:Accession: AH0461  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-357 <KUR>  
A:Cross-references: GB:AL590842; PIDN:CAC93260.1; PID:gl5981707; GSPDB:GN00175  
C:Genetics:  
A:Gene: ugpC  
C:Superfamily: inner membrane protein malk; ATP-binding cassette homology

Query Match 40.2%; Score 51.5; DB 2; Length 357;  
Best Local Similarity 41.4%; Pred. NO. 22;  
Matches 12; Conservative 1; Mismatches 7; Indels 9; Gaps 1;

QY 1 POEIQPOW-----TPEEDQKAREA 20  
DB 263 PLEIQPOWGGRRLLGIRPEHIQQTSA 291

RESULT 13  
95384

A:Title: myb-protein [imported] - Arabidopsis thaliana  
A:Species: Arabidopsis thaliana (mouse-ear cress)  
A:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 17-May-2002  
C:Accession: E85384  
R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring Nature 402, 769-777, 1999  
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.  
A:Reference number: AB5001; MUID:20083488; PMID:10617198  
A:Accession: E85384  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-776 <STO>  
A:Cross-references: GB:NC\_001268; NID:g7270220; PIDN:CAB79990.1; GSPDB:GN00140  
C:Genetics:  
A:Gene: AT4G32730  
A:Map position: 4  
C:Superfamily: unassigned myb DNA-binding repeat proteins; myb DNA-binding repeat homol

Query Match 40.2%; Score 51.5; DB 2; Length 776;  
Best Local Similarity 61.1%; Pred. NO. 50;  
Matches 11; Conservative 2; Mismatches 2; Indels 3; Gaps 1;

QY 8 QWTPEEDQ---KAREAFR 22  
||||||| :| | | :

DB 37 QWTPEDEVLCVERFQ 54

RESULT 14  
T22552

A:Title: hypothetical protein ZK1151.1 - Caenorhabditis elegans  
A:Species: Caenorhabditis elegans  
A:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T22552; T27703  
R:Harris, B.  
submitted to the EMBL Data Library, March 1997

A:Reference number: Z19580  
A:Accession: T22552  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1355 <WIL>  
A:Cross-references: EMBL:Z92788; PIDN:CAB07214.1; GSPDB:GN00019; CESP:ZK1151.1  
A:Experimental source: clone F53B8  
R:Harris, B.  
submitted to the EMBL Data Library, March 1997

A:Reference number: Z20408  
A:Accession: T27703  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1355 <WIL>  
A:Cross-references: EMBL:Z93398; PIDN:CAB07724.1; GSPDB:GN00019; CESP:ZK1151.1  
A:Experimental source: clone ZK1151  
C:Genetics:  
A:Gene: CESP:ZK1151.1  
A:Map position: 1  
A:Introns: 94/3; 124/3; 150/3; 209/3; 248/3; 311/3; 424/3; 734/2; 934/2; 1104/3; 1167

Query Match 39.8%; Score 51; DB 2; Length 1355;  
Best Local Similarity 36.4%; Pred. NO. 1.1e+02;  
Matches 8; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 2 QEIQPOWTPPEEDQKAREAFRR 23

DB 996 KRPPEDWASQYDKMAELLKK 1017

RESULT 15  
S33643

A:Title: transforming protein B-myb - African clawed frog  
N:Alternate names: transforming protein myb1  
C:Species: Xenopus laevis (African clawed frog)  
A:Date: 05-Mar-1994 #sequence\_revision 01-Sep-1995 #text\_change 07-May-1999  
C:Accession: S33643; S27942  
R:Bouwmeester, T.; Gueehmann, S.; El-Baradi, T.; Kalkbrenner, F.; van Wijk, I.; Moell Mech. Dev. 37, 57-68, 1992  
A:Title: Molecular cloning, expression and in vitro functional characterization of Myb  
A:Reference number: S33643; MUID:92297434; PMID:1606020  
A:Accession: S33643  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-733 <BOU>  
A:Cross-references: EMBL:M75870; NID:g214597; PID:g214598  
C:Genetics:  
A:Gene: myb1; B-myb  
C:Function:  
A:Description: transcription regulation; widespread activator of cell cycle genes; re  
C:Superfamily: myb transforming protein; myb DNA-binding repeat homology  
C:Keywords: DNA binding; duplication; nucleus; transcription regulation  
F:26-77/Domain: myb DNA-binding repeat homology <MYB1>  
F:78-129/Domain: myb DNA-binding repeat homology <MYB2>  
F:130-180/Domain: myb DNA-binding repeat homology <MYB3>  
F:372-375/Region: nuclear location signal  
F:528-531/Region: nuclear location signal

Query Match 39.5%; Score 50.5; DB 1; Length 733;  
Best Local Similarity 28.6%; Pred. NO. 64;  
Matches 8; Conservative 8; Mismatches 5; Indels 7; Gaps 1;

Qy 3 EIPQP-----QWTPEDQKAREAFRR 23  
Db 21 DVPEFENRVKVKWTPEDETALKALVKK 48

Search completed: January 2, 2003, 13:45:35  
Job time : 5.43145 secs

This Page Blank (uspto)

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2003, 13:40:48 ; Search time 1.85484 Seconds  
(without alignments)  
514.306 Million cell updates/sec

Title: us-09-818-066-34\_copy\_80\_102  
Perfect score: 128  
Sequence: 1 PQEIPQWTPEDQKAREAFRR 23

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	128	100.0	328	1 VMSA_HPBDDU	P03145 duck hepati
2	128	100.0	365	1 VMSA_HPBDC	P30029 duck hepati
3	121	94.5	366	1 VMSA_HPBDDW	P17195 duck hepati
4	118	92.2	366	1 VMSA_HPBDB	P17194 duck hepati
5	62.5	48.8	335	1 VMSA_HPBHE	P13847 heron hepat
6	54	42.2	390	1 YD3A_SCHPO	Q10274 schizosacch
7	51	39.8	3726	1 ABFI_MOUSE	O61329 mus musculu
8	50.5	39.5	743	1 MYBB_XENLA	P52551 xenopus lae
9	50	39.1	172	1 CATL_HUMAN	P41208 homo sapien
10	50	39.1	624	1 MYB_XENLA	Q08759 xenopus lae
11	50	39.1	986	1 GM13_RAT	O62839 rattus norv
12	50	39.1	1078	1 GYRE_SYNY3	P77966 synecocyst
13	49	38.3	86	1 OAG1_VIBCH	O9kuh2 vibrio chol
14	49	38.3	148	1 NCCX_ALCXX	O44582 alcaligenes
15	49	38.3	754	1 2287_HUMAN	Q9hbc7 homo sapien
16	49	38.3	2225	1 PYRI_MESAU	P08955 mesocricetu
17	48	37.5	172	1 CAT2_HUMAN	Q12798 homo sapien
18	48	37.5	172	1 CATR_MOUSE	P41209 mus musculu
19	48	37.5	459	1 STBL_YEAST	P22465 saccharomyc
20	48	37.5	613	1 CIK5_HUMAN	P22460 homo sapien
21	48	37.5	637	1 RES1_SCHPO	P33520 schizosacch
22	48	37.5	751	1 MYBA_MOUSE	P51960 mus musculu
23	48	37.5	752	1 MYBA_HUMAN	P10243 homo sapien
24	48	37.5	865	1 LOX2_ORYSA	P29250 oryza sativ
25	47.5	37.1	137	1 UCR7_SCHPO	O74533 schizosacch
26	47	36.7	334	1 YNHG_ECOLI	P76193 escherichia
27	47	36.7	571	1 IF2_THERH	P48515 thermus the
28	47	36.7	602	1 CIK5_MOUSE	O61762 mus musculu
29	47	36.7	806	1 TRAI_MAIZE	P08770 zea mays (m
30	47	36.7	839	1 TRAI_MAIZE	P03010 zea mays (m
31	47	36.7	864	1 LOXX_SOYBN	P24095 glycine max
32	47	36.7	943	1 BLIA_CABEL	P51559 caenorhabdi
33	47	36.7	1733	1 VNDA_PRVKA	P33485 pseudorabdi

## RESULT 1

ID	VMSA_HPBDDU	STANDARD;	PRT;	328 AA.
AC	P03145;			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	01-JUL-1993 (Rel. 26, Last sequence update)			
DT	01-FEB-1994 (Rel. 28, Last annotation update)			
DE	Major surface antigen precursor.			
DE	S.			
GN	Duck hepatitis B virus (DHBV).			
OC	Viruses; Retroloid viruses; Hepadnaviridae; Avihepadnavirus.			
OX	NCBI_TaxID=12639;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=84138772; PubMed=6699938;			
RA	Mandart E., Kay A., Galibert F.;			
RT	"Nucleotide sequence of a cloned duck hepatitis B virus genome.			
RT	comparison with woodchuck and human hepatitis B virus sequences.";			
RL	J. Virol. 49:782-792(1984).			
RN	[2]			
RC	SEQUENCE FROM N.A.			
RP	STRAIN=Isolate DHBV FI-6;			
RX	MEDLINE=91045092; PubMed=2235507;			
RA	Mattes F., Tong S., Teubner K., Blum H.E.;			
RT	"Complete nucleotide sequence of a German duck hepatitis B virus.";			
RL	Nucleic Acids Res. 18:6140-6140(1990).			
RN	[3]			
RP	MYRISTOYLATION.			
RX	MEDLINE=91135002; PubMed=1994583;			
RA	Macrae D.R., Bruss V., Ganem D.;			
RT	"Myristylation of a duck hepatitis B virus envelope protein is			
RT	essential for infectivity but not for virus assembly.";			
RL	Virology 181:359-363(1991).			
CC	-1- PTM: MYRISTOYLATION CONTRIBUTES IMPORTANTLY TO DHBV INFECTIVITY.			
CC	IT IS MOST LIKELY REQUIRED FOR AN EARLY STEP OF THE LIFE CYCLE			
CC	INVOLVING THE ENTRY OR UNCOATING OF VIRUS PARTICLES.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL; X12798; CAB57224.1; -			
DR	PIR; A03710; SAVLD.			
DR	PIR; S12845; S12845.			
DR	InterPro: IPR000349; Hepvir_surtag.			
DR	Pfam: PF00695; VMSA; 1.			
KW	Antigen; Myristate; Envelope protein; Lipoprotein.			
FT	PROPEP 1 161			
FT	CHAIN 162 328			
FT	MAJOR SURFACE ANTIGEN.			
FT	LIPID 2 2			
FT	MYRISTATE.			
FT	CARBOHYD 134 134			
FT	N-LINKED (GLCNAC. . .) (POTENTIAL).			
FT	CARBOHYD 260 260			
FT	N-LINKED (GLCNAC. . .) (POTENTIAL).			

P12845 caenorhabdi  
P27708 homo sapien  
P13337 bacterioph  
P72689 synecocyst  
P44687 haemophilus  
P76055 escherichia  
Q9rrf6 deinococcus  
P35263 marburg vir  
Q9p2v4 homo sapien  
Q00872 homo sapien  
Q8yx02 anabaena sp  
P23072 myxococcus

```
SQ SEQUENCE 328 AA; 36230 MW; B2D771241E407456 CRC64;
Query Match 100.0%; Score 128; DB 1; Length 328;
Best Local Similarity 100.0%; Pred. No. 3.4e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 POEIPQOWTPEEDQKAREAFRR 23
DB 80 POEIPQOWTPEEDQKAREAFRR 102

RESULT 2
VMSA_HPBDG STANDARD; PRT; 365 AA.
ID P30029;
AC P30029;
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Major surface antigen precursor.
GN S.
OS Duck hepatitis B virus (strain China) (DHBV).
OC Viruses; Retroid viruses; Hepadnaviridae; Avihepadnavirus.
OX NCBI_TaxID=31510;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91045091; PubMed=2235506;
RA Tong S., Mattes F., Teubner K., Blum H.E.;
RT "Complete nucleotide sequence of a Chinese duck hepatitis B virus.";
RL Nucleic Acids Res. 18:6139-6139(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M21953; AAA45746.1; -
DR PIR; S12842; SAVLWD.
DR InterPro: IPR000349; Hepvir_surfaq.
DR Pfam; PF00695; VMSA; 1.
DR Antigen.
KW PROPEP
FT CHAIN 1 198
FT CARBOHYD 199 365
FT CARBOHYD 297 297
FT CARBOHYD 297 297
SQ SEQUENCE 365 AA; 40511 MW; 67F3A4174CB7D884 CRC64;

Query Match 100.0%; Score 128; DB 1; Length 365;
Best Local Similarity 100.0%; Pred. No. 3.8e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 POEIPQOWTPEEDQKAREAFRR 23
DB 116 POEIPQOWTPEEDQKAREAFRR 138

RESULT 3
VMSA_HPBDW STANDARD; PRT; 366 AA.
ID P17193;
AC P17193;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 01-MAY-1991 (Rel. 18, Last annotation update)
DE Major surface antigen precursor.
GN S.
OS Duck hepatitis B virus (white Shanghai duck isolate S31) (DHBV).
OC Viruses; Retroid viruses; Hepadnaviridae; Avihepadnavirus.
OX NCBI_TaxID=10440;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90085807; PubMed=2596031;
RA Uchida M., Esumi M., Shikata T.;
```

```
RT "Molecular cloning and sequence analysis of duck hepatitis B virus
RL genomes of a new variant isolated from Shanghai ducks.";
CC Virology 173:600-606(1989).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M32991; AAA45752.1; ALT_INIT.
DR PIR; D33746; SAVLWD.
DR InterPro: IPR000349; Hepvir_surfaq.
DR Pfam; PF00695; VMSA; 1.
DR Antigen.
KW PROPEP
FT CHAIN 1 199
FT CARBOHYD 200 366
FT CARBOHYD 170 170
FT CARBOHYD 298 298
FT CARBOHYD 298 298
SQ SEQUENCE 366 AA; 40858 MW; CF60E78B7B2FCD52 CRC64;

Query Match 94.5%; Score 121; DB 1; Length 366;
Best Local Similarity 95.7%; Pred. No. 3.3e-09;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 POEIPQOWTPEEDQKAREAFRR 23
DB 116 POEIPQOWTPEEDQKAREAFRR 138

RESULT 4
VMSA_HPBDG STANDARD; PRT; 366 AA.
ID P17194;
AC P17194;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 01-MAY-1991 (Rel. 18, Last annotation update)
DE Major surface antigen precursor.
GN S.
OS Duck hepatitis B virus (brown Shanghai duck isolate S5) (DHBV).
OC Viruses; Retroid viruses; Hepadnaviridae; Avihepadnavirus.
OX NCBI_TaxID=10439;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90085807; PubMed=2596031;
RA Uchida M., Esumi M., Shikata T.;
RT "Molecular cloning and sequence analysis of duck hepatitis B virus
RL genomes of a new variant isolated from Shanghai ducks.";
CC Virology 173:600-606(1989).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M32990; AAA45755.1; ALT_INIT.
DR PIR; C33746; SAVLBD.
DR InterPro: IPR000349; Hepvir_surfaq.
DR Pfam; PF00695; VMSA; 1.
DR Antigen.
KW PROPEP
FT CHAIN 1 199
FT CARBOHYD 200 366
FT CARBOHYD 170 170
FT CARBOHYD 298 298
SQ SEQUENCE 366 AA; 40897 MW; 5B72879A182EFF38 CRC64;

Query Match 92.2%; Score 118; DB 1; Length 366;
Best Local Similarity 91.3%; Pred. No. 8.4e-09;
```

Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
Brooks K., Brown D., Brown S., Chillingworth T., Church C.M.,  
Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
Skellton J., Simmonds M., Squares R., Squares S., Stevens K.,  
Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
Woodward J., Volkhardt G., Aert R., Robben J., Grynolprez B.,  
Welters J., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,  
Borzyn K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
Goffeau A., Cadieu E., Dreano S., Gloux S., Leleau V., Mottier S.,  
Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsberg S.L.,  
Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
Ripakovski G.V., Ussery D., Barrrell B.G., Nurse P.;  
"The genome sequence of Schizosaccharomyces pombe.";  
Nature 415:871-880(2002).  
-1- SIMILARITY: CONTAINS 2 MYB-LIKE DOMAINS  
---  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
---  
CC EMBL; Z69729; CAA93598.1; --  
CC HSSP; P54274; IBA5.  
DR InterPro; IPR001005; MYB\_DNA\_binding.  
DR Pfam; PF00249; myb\_DNA-binding; 2.  
DR SMART; SM00395; SANT; 2.  
DR PROSITE; PSS0090; MYB\_3; 2.  
KW Hypothetical protein; Repeat.  
FT DNA\_BIND 50 105 MYB 1.  
FT DNA\_BIND 135 189 MYB 2.  
FT SEQUENCE 390 AA; 43922 MW; 0718C86A44CA7DF1 CRC64;  
Query Match 42.2%; Score 54; DB 1; Length 390;  
Best Local Similarity 36.4%; Pred. No. 3.7;  
Matches 12; Conservative 5; Mismatches 4; Indels 12; Gaps 1;  
QY 1 PQETP-----QPQWTFEEDQKAREAF 21  
| | | | : | | | | | | | | | |  
DB 123 POKIPHVGLSKSTRKERTKQFTPEEDRLLEGF 155  
---  
RESULT 7  
-ABFL\_MOUSE  
ID ABFL\_MOUSE STANDARD; PRT; 3726 AA.  
AC Q61329;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Alpha-fetoprotein enhancer binding protein (AT motif-binding factor)  
DE (AT-binding transcription factor 1).  
GN ATBF1.  
DE Mus musculus (Mouse).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BALB/WK X ICR; TISSUE=Brain.

RX MEDLINE=96194902; PubMed=8654949;  
RA Ido A., Miura Y., Watanabe M., Sakai M., Inoue Y., Miki T.,  
RA Hashimoto T., Morinaga T., Nishi S., Tamaoki T.;  
RT "Cloning of the cDNA encoding the mouse ATB1 transcription factor.";  
RL Gene 168:227-231(1996).  
CC -!- FUNCTION: Transcriptional activator that binds to the AT-rich core  
CC sequence of the enhancer element of the AFP gene.  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- SIMILARITY: CONTAINS 4 HOMEBOX DOMAINS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
DR EMBL; D26046; BAA05046.1; -  
DR HSPSP; P20263; LOCPI.  
DR TRANSFAC; T03881; -  
DR MGD; MGI:99948; Atbf1.  
DR InterPro; IPR001356; Homeobox.  
DR InterPro; IPR000822; Znf\_C2H2.  
DR InterPro; IPR003804; Znf\_U1.  
DR Pfam; PF00046; homeobox; 4.  
DR Pfam; PF00096; zf-C2H2; 20.  
DR ProDom; PD000010; Homeobox; 4.  
DR SMART; SM00389; Hox; 4.  
DR SMART; SM00355; Znf\_C2H2; 22.  
DR SMART; SM00451; Znf\_U1; 7.  
DR PROSITE; PS00027; HOMEBOX\_1; 2.  
DR PROSITE; PS00071; HOMEBOX\_2; 4.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 15.  
DR PROSITE; PS00157; ZINC\_FINGER\_C2H2\_2; 9.  
DR PROSITE; PS00157; ZINC\_FINGER\_C2H2\_2; 9.  
KW Transcription regulation; Activator; Zinc-finger; Metal-binding;  
KW DNA-binding; Homeobox; Nuclear protein; Repeat.  
FT ZN\_FING 79 103  
FT C2H2-TYPE.  
FT ZN\_FING 282 305  
FT C2H2-TYPE.  
FT ZN\_FING 641 664  
FT C2H2-TYPE.  
FT ZN\_FING 672 695  
FT C2H2-TYPE.  
FT ZN\_FING 727 751  
FT C2H2-TYPE.  
FT ZN\_FING 805 829  
FT C2H2-TYPE (ATYPICAL).  
FT ZN\_FING 946 969  
FT C2H2-TYPE (DEGENERATE).  
FT ZN\_FING 985 1009  
FT C2H2-TYPE (ATYPICAL).  
FT ZN\_FING 1041 1065  
FT C2H2-TYPE (ATYPICAL).  
FT ZN\_FING 1089 1113  
FT C2H2-TYPE (ATYPICAL).  
FT ZN\_FING 1233 1256  
FT C2H2-TYPE (ATYPICAL).  
FT ZN\_FING 1262 1285  
FT C2H2-TYPE.  
FT ZN\_FING 1370 1395  
FT C2H2-TYPE.  
FT ZN\_FING 1411 1433  
FT C2H2-TYPE.  
FT ZN\_FING 1439 1462  
FT C2H2-TYPE.  
FT ZN\_FING 1555 1579  
FT C2H2-TYPE.  
FT ZN\_FING 1606 1630  
FT C2H2-TYPE.  
FT ZN\_FING 1980 2013  
FT C2H2-TYPE.  
FT ZN\_FING 2152 2211  
FT HOMEBOX 1.  
FT ZN\_FING 2249 2308  
FT HOMEBOX 2.  
FT ZN\_FING 2335 2358  
FT C2H2-TYPE (ATYPICAL).  
FT ZN\_FING 2539 2561  
FT C2H2-TYPE.  
FT ZN\_FING 2650 2709  
FT HOMEBOX 3.  
FT ZN\_FING 2720 2743  
FT C2H2-TYPE.  
FT ZN\_FING 2952 3011  
FT HOMEBOX 4.  
FT ZN\_FING 3032 3056  
FT C2H2-TYPE.  
FT ZN\_FING 3552 3576  
FT C2H2-TYPE.  
FT ZN\_FING 461 491  
FT POLY-GLU.  
FT ZN\_FING 491 785  
FT POLY-ALA.  
FT ZN\_FING 1314 1317  
FT POLY-ALA.  
FT ZN\_FING 1734 1748  
FT POLY-GLN.  
FT ZN\_FING 1794 1799  
FT POLY-GLN.  
FT ZN\_FING 1856 1863  
FT POLY-GLN.  
FT ZN\_FING 2044 2059  
FT POLY-PRO.  
FT ZN\_FING 2405 2408  
FT POLY-ALA.  
FT ZN\_FING 3216 3220  
FT POLY-PRO.

FT DOMAIN 3380 3409 POLY-GLN.  
FT DOMAIN 3412 3420 POLY-GLN.  
FT DOMAIN 3534 3550 POLY-GLY.  
FT DOMAIN 3620 3623 POLY-PRO.  
FT DOMAIN 3659 3662 POLY-SER.  
SQ SEQUENCE 3726 AA; 406567 MW; 915ACBE588A72C98 CRC64;  
Query Match 39.8%; Score 51; DB 1; Length 3726;  
Best Local Similarity 44.4%; Pred. No. 1.1e+02;  
Matches 8; Conservative. 6; Mismatches 4; Indels 0; Gaps 0;  
QY 2 QEIPOQWTPEDQKARE 19  
DB 3223 QQIPAPQLTPQQQRKDKD 3240  
RESULT 8  
MYBB\_XENLA  
ID MYBB\_XENLA STANDARD; PRT; 743 AA.  
AC P52551;  
DT 01-OCT-1996 (Rel. 34; Created)  
DT 30-MAY-2000 (Rel. 39; Last sequence update)  
DE 16-OCT-2001 (Rel. 40; Last annotation update)  
DE Myb-related protein B (B-Myb) (Myb-related protein 1) (XMYB1).  
GN MYBL2 OR BMYB OR MYBL.  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
OC Xenopodinae; Xenopus.  
OX NCBI\_TaxID=8355;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92297434; PubMed=1606020;  
RA Bouwmeester T., Guelmann S., El-Baradi T., Kalkbrenner F.,  
RA van Wijk I., Moelling K., Pieler T.;  
RT "Molecular cloning, expression and in vitro functional  
RT characterization of Myb-related proteins in Xenopus.";  
RL Mech. Dev. 37:57-68(1992).  
RN [2]  
RP REVISIONS TO C-TERMINUS.  
RA Humbert-Lan G., Pieler T.;  
RT "Regulation of DNA-binding activity and nuclear transport of B-Myb in  
RT Xenopus oocytes.";  
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- DEVELOPMENTAL STAGE: PRESENT THROUGHOUT OOCYTESIS AND EARLY  
CC XENOPUS EMBRYOGENESIS; IN ADULT TISSUE IT IS PRIMARILY DETECTED IN  
CC BLOOD.  
CC -!- SIMILARITY: CONTAINS 3 MYB-LIKE DOMAINS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M75870; AAC98701.1; -  
DR HSPSP; Q03237; IASJ  
DR InterPro; IPR001005; Myb\_DNA\_binding.  
DR Pfam; PF00249; myb\_DNA-binding; 3.  
DR SMART; SM00395; SANT; 3.  
DR PROSITE; PS00037; MYB\_1; 3.  
DR PROSITE; PS00334; MYB\_2; 3.  
DR PROSITE; PS00090; MYB\_3; 3.  
DR Transcription regulation; Nuclear protein; DNA-binding; Repeat.  
FT ZN\_FING 26 77  
FT MYB 1.  
FT ZN\_FING 78 129  
FT MYB 2.  
FT ZN\_FING 130 180  
FT MYB 3.  
SQ SEQUENCE 743 AA; 82909 MW; 5FD1D678BB24409B CRC64;  
Query Match 39.5%; Score 50.5; DB 1; Length 743;



DR	'SMART; SMO0054; EFH; 4.
DR	PROSITE; PS00018; EF_HAND; 2.
FT	Calcium-binding; Repeat; Cell division; Mitosis.
FT	CA_BIND 41 52 EF-HAND 1 (PROBABLE).
FT	DOMAIN 77 88 ANCESTRAL CALCIUM SITE 2 (POTENTIAL).
FT	DOMAIN 114 125 ANCESTRAL CALCIUM SITE 3 (POTENTIAL).
FT	CA_BIND 150 161 EF-HAND 4 (PROBABLE).
SQ	SEQUENCE 172 AA; 19738 MW; 59CFD706AD7011B5 CRC64;
Query Match . 39.1%; Score 50; DB 1; Length 172;	
Best Local Similarity 52.9%; Pred. No. 5.4;	
Matches	9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
QY	5 PQQWTPPEEDQKAREAF 21
ID	I : I : I : I : I :
DB	21 PKPELTTEEQKEIREAF 37
RESULT 10	
MYB_XENLA	STANDARD; PRT; 624 AA.
ID	MYB_XENLA STANDARD; PRT; 624 AA.
AC	O08759;
DT	01-FEB-1995 (Rel. 31, Created)
DT	01-FEB-1995 (Rel. 31, Last sequence update)
DT	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Myb protein.
GN	MYB.
OS	Xenopus laevis (African clawed frog).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC	Xenopodinae; Xenopus.
OX	NCBI_TaxID=8355;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=94151023; PubMed=7509053;
RA	Anaravadi L., King M.W.;
RT	"Characterization and expression of the Xenopus c-Myc homolog.";
RL	Oncogene 9:971-974(1994).
CC	-I- FUNCTION: TRANSCRIPTIONAL ACTIVATOR; DNA-BINDING PROTEIN THAT SPECIFICALLY RECOGNIZE THE SEQUENCE 5'-YAAC(G/T)G-3'. PLAYS AN IMPORTANT ROLE IN THE CONTROL OF PROLIFERATION AND DIFFERENTIATION OF HEMATOPOIETIC PROGENITOR CELLS.
CC	-I- SUBCELLULAR LOCATION: Nuclear.
CC	-I- DOMAIN: COMPRISED OF 3 DOMAINS; AN N-TERMINAL DNA-BINDING DOMAIN, A CENTRALLY LOCATED TRANSCRIPTIONAL ACTIVATION DOMAIN AND A C-TERMINAL DOMAIN INVOLVED IN TRANSCRIPTIONAL REPRESSON.
CC	-I- SIMILARITY: CONTAINS 3 MYB-LIKE DOMAINS.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outpost - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; L22741; AAC38011.1; ..
DR	HSSP; P06876; IMEG.
DR	InterPro; IPRO01005; Myb_DNA_binding.
DR	Pfam; PF00249; myb_DNA-binding; 3.
DR	'SMART; SMO0395; SANT; 3.
DR	PROSITE; PS00037; MYB_1; 3.
DR	PROSITE; PS00334; MYB_2; 3.
DR	PROSITE; PS50030; MYB_3; 3.
KW	Transcription regulation; Activator; Nuclear protein; DNA-binding; Repeat.
FT	DNA_BIND 32 83 MYB 1.
FT	DNA_BIND 84 135 MYB 2.
FT	DNA_BIND 136 186 MYB 3.
FT	DOMAIN 264 314 TRANSRIPTIONAL ACTIVATION DOMAIN (BY SIMILARITY). NEGATIVE REGULATORY DOMAIN (BY SIMILARITY).
FT	DOMAIN 315 449

```
SQ SEQUENCE 624 AA; 72112 MW; B8F51A2BBA72E70B CRC64;
Query Match 39.1%; Score 50; DB 1; Length 624;
Best Local Similarity 42.9%; Pred. No. 21;
Matches 9; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 3 EIQPQWTPEDQKAREAFRR 23
: : ||| :
Db 138 EVKSSWTEEDRTIYEAHR 158

RESULT 11
GM13_RAT STANDARD; PRT; 986 AA.
ID Q62839;
AC Q62839;
DT 01-NOV-1997 (Rel. 35, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Cis-golgi matrix protein GM130.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE OF 76-986 FROM N.A., AND CHARACTERIZATION.
RC STRAIN-Sprague-Dawley;
RX MEDLINE=96134022; PubMed=8557739;
RA Nakamura N., Rabouille C., Watson R., Nilsson T., Hui N.,
RA Slusarewicz P., Kreis T.E., Warren G.;
PT "Characterization of a cis-Golgi matrix protein, GM130."
RL J. Cell Biol. 131:1715-1726(1995).
RN [2]
RP REVISIONS, AND CHARACTERIZATION.
RC STRAIN-Sprague-Dawley;
RX MEDLINE=97294388; PubMed=9150144;
RA Nakamura N., Lowe M., Levine T.P., Rabouille C., Warren G.;
RT "The vesicle docking protein p115 binds GM130, a cis-Golgi matrix
RT protein, in a mitotically regulated manner."
RL Cell 89:445-455(1997).
RN [3]
RP PHOSPHORYLATION OF SER-25.
RX MEDLINE=98424247; PubMed=9753325;
RA Lowe M., Rabouille C., Nakamura N., Watson R., Jackman M., Jamsa E.,
RA Rahman D., Pappin D.J., Warren G.;
RT "Cdc2 kinase directly phosphorylates the cis-Golgi matrix protein
RT GM130 and is required for Golgi fragmentation in mitosis."
RL Cell 94:783-793(1998).
CC -1- FUNCTION: PROBABLY HAS A ROLE IN MAINTAINING CIS-GOLGI STRUCTURE.
CC -1- SUBUNIT: PART OF A LARGER OLIGOMERIC COMPLEX. INTERACTS WITH P115.
CC -1- SUBCELLULAR LOCATION: PERIPHERAL CYTOPLASMIC PROTEIN THAT IS
CC TIGHTLY BOUND TO MEMBRANES OF THE CIS-GOLGI NETWORK. PRESENT ON
CC THE CYTOPLASMIC SIDE OF THE MEMBRANE.
CC -1- DOMAIN: EXTENDED ROD-LIKE PROTEIN WITH COILED-COIL DOMAINS.
CC -1- SIMILARITY: HIGH, TO HUMAN GOLGIN 95.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
```

```

Query Match      38.3%; Score 49; DB 1; Length 86;
Best Local Similarity 44.4%; Pred. No. 3.5;
Matches      8; Conservative    5; Mismatches    5; Indels    0; Gaps    0;

QY   1 PQEIPQPQTPEEDOKAR 18
      |||::||::||::||
DB    41 PQEVPEAAATPKSKVKV 58

RESULT 14
NCXX_ALCXX
ID NCXX_ALCXX STANDARD; PRT; 148 AA.
AC Q44582;
DT DT 15-DEC-1998 (Rel. 37, Created)
DT DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE DE Nickel-cobalt-cadmium resistance protein nccx.
GN NCXX.
OS Alcaligenes xylosoxydans xylosoxydans (Achromobacter xylosoxidans).
OG Plasmid pWOM9.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Achromobacter.
OX NCBI_TaxID=515;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=31A;
RX MEDLINE=95050278; PubMed=7961470;
RT Schmidt T., Schlegel H.G.;
RA "Combined nickel-cobalt-cadmium resistance encoded by the ncc locus
of Alcaligenes xylosoxidans 31A."
RL J. Bacteriol. 176:7045-7054(1994).
CC -1- FUNCTION: COMPONENT OF THE NCC CATION-EFFLUX SYSTEM THAT CONFERS
CC RESISTANCE TO NICKEL, COBALT AND CADMIUM. MAY BE INVOLVED IN THE
CC REGULATION OF NCC.
CC -1- SIMILARITY: TO A.EUTROPHUS CNRR.
-----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcements
CC or send an email to license@isb-sib.ch).
-----
EMBL: L31363; AAA65102.1;
DR Plasmid; Nickel; Cobalt; Cadmium.
SQ SEQUENCE 148 AA; 16685 MW; BC9F9GF54271B830 CRC64;

Query Match      38.3%; Score 49; DB 1; Length 148;
Best Local Similarity 50.0%; Pred. No. 6.2;
Matches      3; Conservative    3; Mismatches    5; Indels    0; Gaps    0;

QY   7 PQWTPEEQKAREAFR 22
      |||::||::||
DB    92 PQWSPEVEATEPREVER 107

RESULT 15
Z287_HUMAN
ID Z287_HUMAN STANDARD; PRT; 754 AA.
AC Q9HB7;
DT DT 15-JUN-2002 (Rel. 41, Created)
DT DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE DE Zinc finger protein ZNF287.
GN ZNF287.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.

```

RA Romero-Pastrana F., Srivastava A.K.;  
RL "Deletion of a novel zinc finger gene in Smith-Magenis syndrome.";  
CC Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.  
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).  
CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-  
CC FINGER PROTEINS.  
CC -!- SIMILARITY: CONTAINS 1 SCAN BOX.  
CC -!- SIMILARITY: CONTAINS 1 KRAB DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AF217227; AGO9702.1; -;  
CC HSP; P07248; IPAA.  
CC Genew; HGNC:13502; ZNF287.  
CC InterPro; IPR001909; KRAB.  
CC InterPro; IPR003309; Treg\_SCAN.  
CC InterPro; IPR000822; Znf\_C2H2.  
CC Pfam; PF00096; zf-C2H2; 14.  
CC Pfam; PF01352; KRAB; 1.  
CC Pfam; PF02023; SCAN; 1.  
CC PRINTS; PR00048; ZINCFINGER.  
CC ProDom; PD000003; Znf\_C2H2; 12.  
CC SMART; SM00349; KRAB; 1.  
CC SMART; SM00431; LER; 1.  
CC SMART; SM00355; Znf\_C2H2; 14.  
CC PROSITE; PS0805; KRAB; 1.  
CC PROSITE; PS0804; SCAN\_BOX; 1.  
CC PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 14.  
CC PROSITE; PS0157; ZINC\_FINGER\_C2H2\_2; 14.  
KW Transcription regulation; Zinc-finger; Metal-binding; Nuclear protein;  
KW DNA-binding; Repeat.  
FT DOMAIN 42 124 SCAN BOX.  
FT DOMAIN 163 231 KRAB.  
FT DOMAIN 361 747 ZINC FINGERS.  
FT ZN\_FING 361 383 C2H2-TYPE.  
FT ZN\_FING 389 411 C2H2-TYPE.  
FT ZN\_FING 417 439 C2H2-TYPE.  
FT ZN\_FING 445 467 C2H2-TYPE.  
FT ZN\_FING 473 495 C2H2-TYPE.  
FT ZN\_FING 501 523 C2H2-TYPE.  
FT ZN\_FING 529 551 C2H2-TYPE.  
FT ZN\_FING 557 579 C2H2-TYPE.  
FT ZN\_FING 585 607 C2H2-TYPE.  
FT ZN\_FING 613 635 C2H2-TYPE.  
FT ZN\_FING 641 663 C2H2-TYPE.  
FT ZN\_FING 669 691 C2H2-TYPE.  
FT ZN\_FING 697 719 C2H2-TYPE.  
FT ZN\_FING 725 747 C2H2-TYPE.  
SQ SEQUENCE 754 AA; 87564 MW; AICCFIF5180601C5 CRC64;

Query Match 38.3%; Score 49; DB 1; Length 754;  
Best Local Similarity 45.5%; Pred. No. 36;  
Matches 10; Conservative 3; Mismatches 9; Indels 0; Gaps 0;  
Qy 1 POEIPQPWTEEDQKAREAFR 22  
||| : |||  
Db 129 PQNSTLSQDTPEEDPRGKHAQ 150

Search completed: January 2, 2003, 13:44:52  
Job time : 3.85484 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2003, 13:42:28 ; Search time 6.58468 Seconds  
(without alignments)  
719.714 Million cell updates/sec

Title: US-09-818-066-34\_COPY\_80\_102  
Perfect score: 128  
Sequence: 1 PQEIPQQTPEEDQKAREAFRR 23

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_21.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phage.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_virus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	128	100.0	328	12	Q8QX1 duck hepati
2	128	100.0	329	12	Q92935 duck hepati
3	128	100.0	329	12	Q91HP5 duck hepati
4	128	100.0	364	12	Q66399 duck hepati
5	123	96.1	329	12	Q9WFA3 snow goose
6	123	96.1	329	12	Q9WFA6 snow goose
7	123	96.1	329	12	Q9WFA9 snow goose
8	123	96.1	329	12	Q9WFB3 snow goose
9	123	96.1	329	12	Q9WFB6 snow goose
10	121	94.5	330	12	Q72885 duck hepati
11	118	92.2	330	12	Q66405 duck hepati
12	118	92.2	366	12	Q66404 duck hepati
13	77	60.2	327	12	Q67852 duck hepati
14	57	44.5	104	11	Q9DD24 mus musculus
15	56	43.8	311	2	Q93E30 salmonella
16	56	43.8	311	16	Q8ZP88 salmonella

17	56	43.8	311	16	Q8Z783 salmonella
18	55	43.0	566	10	Q9M596 papaver rho
19	54	42.2	282	3	P78793 schizosacch
20	54	42.2	337	12	Q8UYV0 stork hepat
21	54	42.2	337	12	Q8UYX8 stork hepat
22	54	42.2	337	12	Q8UYX6 stork hepat
23	54	42.2	337	12	Q8UYX4 stork hepat
24	54	42.2	1003	10	Q94BS6 nicotiana t
25	54	42.2	1112	17	Q9UYR7 pyrococcus
26	53	41.4	540	4	Q96M05 homo sapien
27	53	41.4	847	10	Q9LV31 arabidopsis
28	52	40.6	265	2	Q93M36 corynebacte
29	52	40.6	320	10	Q9XHV0 arabidopsis
30	52	40.6	744	16	Q8UJ85 agrobacteri
31	51.5	40.2	357	16	Q8ZAK8 yersinia pe
32	51.5	40.2	776	10	Q9LDX5 arabidopsis
33	51.5	40.2	776	10	Q9S7G7 arabidopsis
34	51	39.8	246	16	Q92KL8 rhizobium m
35	51	39.8	254	10	Q9LD63 adiantum ra
36	51	39.8	319	10	Q9SPG6 arabidopsis
37	51	39.8	1376	5	O18290 caenorhabdi
38	50.5	39.5	481	5	Q95W90 trypanosoma
39	50	39.1	44	10	Q8S422 zea mays (m
40	50	39.1	110	6	Q29204 sus scrofa
41	50	39.1	295	10	Q64399 arabidopsis
42	50	39.1	321	10	Q9LDI5 arabidopsis
43	50	39.1	350	12	Q8QMX3 cowpox viru
44	50	39.1	496	10	Q9SR26 arabidopsis
45	50	39.1	505	10	Q9M652 arabidopsis

## ALIGNMENTS

## RESULT 1

Q8QX1 ID Q8QX1 PRELIMINARY; PRT; 328 AA.  
AC Q8QX1;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DE Presurface protein.  
OS Duck hepatitis B virus (DHBV).  
OC Viruses; Retroid viruses; Hepadnaviridae; Avihepadnavirus.  
OX NCBI\_TaxID=12639;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=INDIANA;  
RA Wang C.-Y.J., Glambrone J.J., Dormitorio T.V.;  
RT "The complete sequence of Duck Hepatitis B virus Indiana isolate."  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF493986; AAM11781.1;  
SQ SEQUENCE 328 AA; 36146 MW; BE6DIC9E73FA1556 CRC64;

Query Match 100.0%; Score 128; DB 12; Length 328;  
Best Local Similarity 100.0%; Pred. No. 1.3e-09;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PQEIPQQTPEEDQKAREAFRR 23

Db 80 PQEIPQQTPEEDQKAREAFRR 102

## RESULT 2

Q92935 ID Q92935 PRELIMINARY; PRT; 329 AA.  
AC Q92935;  
DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
DE Pres antigen.  
OS Duck hepatitis B virus (DHBV).  
OC Viruses; Retroid viruses; Hepadnaviridae; Avihepadnavirus.

```
OX NCBI_TaxID=12639;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=ALBERTA;
RA Fischer K.P., Stickney J., Tipples G.A., Tyrrell D.L.J.;
RT "Cloning, sequencing and sequence comparison of a Canadian isolate of
RT duck hepatitis B virus.";
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF047045; AAC06355.1; -.
DR InterPro; IPR000349; Hepvir_surfa.
DR Pfam; PF00695; VNSA; 1.
SQ SEQUENCE 329 AA; 36361 MW; 45E4ACAFCA995147 CRC64;

Query Match 100.0%; Score 128; DB 12; Length 329;
Best Local Similarity 100.0%; Pred. No. 1.3e-09;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 POEIPQWTPEDQKAREAFRR 23
Db 81 POEIPQWTPEDQKAREAFRR 103

RESULT 3
Q91HP5
ID Q91HP5 PRELIMINARY; PRT; 329 AA.
AC Q91HP5;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Pres protein.
OS Duck hepatitis B virus (DHBV).
OC Viruses; Retroid viruses; Hepadnaviridae; Avihepadnavirus.
OX NCBI_TaxID=12639;
RN [1]
RP SEQUENCE FROM N.A.
RA Hu J., Tang N., Huang A.;
RT "Sequence Analysis of a Cloned Duck Hepatitis B Virus Genome from
RT Shongqing Brown Duck.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF040406; AAK85437.1; -.
DR InterPro; IPR000349; Hepvir_surfa.
DR Pfam; PF00695; VNSA; 1.
SQ SEQUENCE 329 AA; 36436 MW; FIDFE48192CE9F97 CRC64;

Query Match 100.0%; Score 128; DB 12; Length 329;
Best Local Similarity 100.0%; Pred. No. 1.3e-09;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 POEIPQWTPEDQKAREAFRR 23
Db 80 POEIPQWTPEDQKAREAFRR 102

RESULT 4
Q96399
ID Q96399 PRELIMINARY; PRT; 364 AA.
AC Q96399;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Surface antigen.
OS Duck hepatitis B virus (DHBV).
OC Viruses; Retroid viruses; Hepadnaviridae; Avihepadnavirus.
OX NCBI_TaxID=12639;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=INDIAN;
RA Munshi A., Panda S.K.;
RT "Cloning sequencing and sequence comparison of the Indian isolate.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; X74623; CAA52699.1; -.
DR InterPro; IPR000349; Hepvir_surfa.
DR Pfam; PF00695; VNSA; 1.
RP SEQUENCE FROM N.A.
```

```
SQ SEQUENCE 364 AA; 40385 MW; E2E27FB4E4775C19 CRC64;

Query Match 100.0%; Score 128; DB 12; Length 364;
Best Local Similarity 100.0%; Pred. No. 1.4e-09;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 POEIPQWTPEDQKAREAFRR 23
Db 116 POEIPQWTPEDQKAREAFRR 138

RESULT 5
Q9WFA3
ID Q9WFA3 PRELIMINARY; PRT; 329 AA.
AC Q9WFA3;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Pres antigen.
OS snow goose hepatitis B virus.
OC Viruses; Retroid viruses; Hepadnaviridae; Avihepadnavirus.
OX NCBI_TaxID=89623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SGHBV1-13;
RX MEDLINE=99420377; PubMed=10489339;
RA Chang S.F., Netter H.J., Bruns M., Schneider R., Frolich K., Will H.;
RT "A new avian hepadnavirus infecting snow geese (Anser caerulescens)
RT produces a significant fraction of virions containing single-stranded
RT DNA.";
RL Virology 262:39-54(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SGHBV1-13;
RA Chang S.-F., Netter H.J., Bruns M., Schneider R., Froelich K.,
RA Will H.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF110996; AAD21982.1; -.
DR InterPro; IPR000349; Hepvir_surfa.
DR Pfam; PF00695; VNSA; 1.
SQ SEQUENCE 329 AA; 36657 MW; 12B2DC5E887FC420 CRC64;

Query Match 96.1%; Score 123; DB 12; Length 329;
Best Local Similarity 95.7%; Pred. No. 6.1e-09;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 POEIPQWTPEDQKAREAFRR 23
Db 80 POEIPQWTPEDQKAREAFRR 102

RESULT 6
Q9WFA6
ID Q9WFA6 PRELIMINARY; PRT; 329 AA.
AC Q9WFA6;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Pres antigen.
OS snow goose hepatitis B virus.
OC Viruses; Retroid viruses; Hepadnaviridae; Avihepadnavirus.
OX NCBI_TaxID=89623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SGHBV1-15;
RX MEDLINE=99420377; PubMed=10489339;
RA Chang S.F., Netter H.J., Bruns M., Schneider R., Frolich K., Will H.;
RT "A new avian hepadnavirus infecting snow geese (Anser caerulescens)
RT produces a significant fraction of virions containing single-stranded
RT DNA.";
RL Virology 262:39-54(1999).
RN [2]
RP SEQUENCE FROM N.A.
```

RC STRAIN-SGHBV1-15;  
RA Chang S.-F., Netter H.J., Bruns M., Schneider R., Froelich K., Will H.;  
RL "A new avian hepadnavirus infecting snow geese (Anser caerulescens)  
DR EMBL; AF110997; AAD21987.1; Last sequence update)  
DR InterPro; IPR000349; Hepvir\_surfaag.  
DR Pfam; PF00695; VMSA; 1.  
SQ SEQUENCE 329 AA; 36586 MW; 7C1928C4C0E87466 CRC64;

Query Match 96.1%; Score 123; DB 12; Length 329;  
Best Local Similarity 95.7%; Pred. No. 6.1e-09;  
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PQEIPQQTPEEDQKAREAFRR 23  
Db 80 PQEIPQPHWTPEEDQKAREAFRR 102

RESULT 7  
Q9WFA9 PRELIMINARY; PRT; 329 AA.  
AC Q9WFA9;  
DT 01-NOV-1999 (Tremblrel. 12, Created)  
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)  
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
DE Pres antigen.  
OS snow goose hepatitis B virus.  
OC Viruses; Retroid viruses; Hepadnaviridae; Avihepadnavirus.  
OX NCBI\_TaxID=89623;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SGHBV1-19;  
RX MEDLINE-99420377; PubMed-10489339;  
RA Chang S.-F., Netter H.J., Bruns M., Schneider R., Froelich K., Will H.;  
RL "A new avian hepadnavirus infecting snow geese (Anser caerulescens)  
RT produces a significant fraction of virions containing single-stranded  
RT DNA.";  
RL Virology 262:39-54 (1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SGHBV1-19;  
RX MEDLINE-99420377; PubMed-10489339;  
RA Chang S.-F., Netter H.J., Bruns M., Schneider R., Froelich K., Will H.;  
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF110998; AAD21992.1; Last sequence update)  
DR InterPro; IPR000349; Hepvir\_surfaag.  
DR Pfam; PF00695; VMSA; 1.  
SQ SEQUENCE 329 AA; 36645 MW; D896E4B2979F201E CRC64;

Query Match 96.1%; Score 123; DB 12; Length 329;  
Best Local Similarity 95.7%; Pred. No. 6.1e-09;  
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PQEIPQQTPEEDQKAREAFRR 23  
Db 80 PQEIPQPHWTPEEDQKAREAFRR 102

RESULT 8  
Q9WFB3 PRELIMINARY; PRT; 329 AA.  
AC Q9WFB3;  
DT 01-NOV-1999 (Tremblrel. 12, Created)  
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)  
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
DE Pres antigen.  
OS snow goose hepatitis B virus.  
OC Viruses; Retroid viruses; Hepadnaviridae; Avihepadnavirus.  
OX NCBI\_TaxID=89623;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SGHBV1-7;  
RX MEDLINE-99420377; PubMed-10489339;

RA Chang S.-F., Netter H.J., Bruns M., Schneider R., Froelich K., Will H.;  
RT "A new avian hepadnavirus infecting snow geese (Anser caerulescens)  
RT produces a significant fraction of virions containing single-stranded  
RT DNA.";  
RL Virology 262:39-54 (1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SGHBV1-7;  
RA Chang S.-F., Netter H.J., Bruns M., Schneider R., Froelich K.,  
RA Will H.;  
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF110999; AAD21997.1; Last sequence update)  
DR InterPro; IPR000349; Hepvir\_surfaag.  
DR Pfam; PF00695; VMSA; 1.  
SQ SEQUENCE 329 AA; 36602 MW; 46DAB552978B0F27 CRC64;

Query Match 96.1%; Score 123; DB 12; Length 329;  
Best Local Similarity 95.7%; Pred. No. 6.1e-09;  
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PQEIPQQTPEEDQKAREAFRR 23  
Db 80 PQEIPQPHWTPEEDQKAREAFRR 102

RESULT 9  
Q9WFB6 PRELIMINARY; PRT; 329 AA.  
AC Q9WFB6;  
DT 01-NOV-1999 (Tremblrel. 12, Created)  
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)  
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
DE Pres antigen.  
OS snow goose hepatitis B virus.  
OC Viruses; Retroid viruses; Hepadnaviridae; Avihepadnavirus.  
OX NCBI\_TaxID=89623;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SGHBV1-9;  
RX MEDLINE-99420377; PubMed-10489339;  
RA Chang S.-F., Netter H.J., Bruns M., Schneider R., Froelich K., Will H.;  
RL "A new avian hepadnavirus infecting snow geese (Anser caerulescens)  
RT produces a significant fraction of virions containing single-stranded  
RT DNA.";  
RL Virology 262:39-54 (1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SGHBV1-9;  
RA Chang S.-F., Netter H.J., Bruns M., Schneider R., Froelich K.,  
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF111000; AAD22002.1; Last sequence update)  
DR InterPro; IPR000349; Hepvir\_surfaag.  
DR Pfam; PF00695; VMSA; 1.  
SQ SEQUENCE 329 AA; 36486 MW; 6A85C3BEEF3FA526 CRC64;

Query Match 96.1%; Score 123; DB 12; Length 329;  
Best Local Similarity 95.7%; Pred. No. 6.1e-09;  
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PQEIPQQTPEEDQKAREAFRR 23  
Db 80 PQEIPQPHWTPEEDQKAREAFRR 102

RESULT 10  
O72885 PRELIMINARY; PRT; 330 AA.  
ID O72885;  
AC O72885;  
DT 01-AUG-1998 (Tremblrel. 07, Created)  
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)  
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
DE Surface protein.

OS Duck hepatitis B virus (DHBV).  
OC Viruses; Retroid viruses; Hepadnaviridae; Avihepadnavirus.  
OX NCBI\_TaxID=12639;  
[1]  
RN SEQUENCE FROM N.A.  
RC STRAIN-AUSTRALIAN DHBV;  
RA Tong S., Mattes F., Blum H.E., Fernholz D., Schneider R., Will H.;  
RT "Complete nucleotide sequence of a Chinese Hepatitis B virus";  
RL Submitted (JUN-1991) to the EMBL/GenBank/DBJ databases.  
RA Triyatni M.;  
DR EMBL; X60213; CAA42770.1; -;  
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
[2]  
RN SEQUENCE FROM N.A.  
RC STRAIN-AUSTRALIAN DHBV;  
RX MEDLINE=21102973; PubMed=11161276;  
RA Triyatni M., Ey P.L., Tran T., Le Mire M., Qiao M., Burrell C.J.,  
RA Jilbert A.R.;  
RT "Sequence comparison of an Australian duck hepatitis B virus strain  
with other avian hepadnaviruses";  
RL J. Gen. Virol. 82:373-378(2001).  
DR EMBL; AJ006350; CAA06988.1; -;  
DR InterPro; IPR000349; Hepvir\_surfa.  
DR Pfam; PF00695; VMSA; 1.  
SQ SEQUENCE 330 AA; 36789 MW; EDED4F42373ADA99 CRC64;

Query Match 94.5%; Score 121; DB 12; Length 330;  
Best Local Similarity 95.7%; Pred. No. 1.1e-08;  
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 POEIPQQTWEEDQKAREAFRR 23  
||| ||||||||||||||||  
Db 80 PQEPPQQTWEEDQKAREAFRR 102

## RESULT 11

Q66405 ID Q66405 PRELIMINARY; PRT; 330 AA.  
AC Q66405;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Pre-S protein.  
OS Duck hepatitis B virus (DHBV).  
OC Viruses; Retroid viruses; Hepadnaviridae; Avihepadnavirus.  
OX NCBI\_TaxID=12639;  
[1]  
RN SEQUENCE FROM N.A.  
RC STRAIN-DHBVCA34;  
RA Tong S., Mattes F., Blum H.E., Fernholz D., Schneider R., Will H.;  
RT "Complete nucleotide sequence of a Chinese Hepatitis B virus";  
RL Submitted (JUN-1991) to the EMBL/GenBank/DBJ databases.  
RA EMBL; X60213; CAA42771.1; -;  
DR InterPro; IPR000349; Hepvir\_surfa.  
DR Pfam; PF00695; VMSA; 1.  
SQ SEQUENCE 330 AA; 36959 MW; 7CE142013BB8D9F4 CRC64;

Query Match 92.2%; Score 118; DB 12; Length 330;  
Best Local Similarity 91.3%; Pred. No. 2.9e-08;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 POEIPQQTWEEDQKAREAFRR 23  
||| ||||||||||||||||  
Db 80 PHETPQQTWEEDQKAREAFRR 102

## RESULT 12

Q66404 ID Q66404 PRELIMINARY; PRT; 366 AA.  
AC Q66404;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Pre-S/S.  
OS Duck hepatitis B virus (DHBV).  
OC Viruses; Retroid viruses; Hepadnaviridae; Avihepadnavirus.  
OX NCBI\_TaxID=12639;

[1]  
RN SEQUENCE FROM N.A.  
RC STRAIN-DHBVCA34;  
RA Tong S., Mattes F., Blum H.E., Fernholz D., Schneider R., Will H.;  
RT "Complete nucleotide sequence of a Chinese Hepatitis B virus";  
RL Submitted (JUN-1991) to the EMBL/GenBank/DBJ databases.  
RA EMBL; X60213; CAA42770.1; -;  
DR InterPro; IPR000349; Hepvir\_surfa.  
DR Pfam; PF00695; VMSA; 1.  
SQ SEQUENCE 366 AA; 41057 MW; FDF3616EBC39629D CRC64;

Query Match 92.2%; Score 118; DB 12; Length 366;  
Best Local Similarity 91.3%; Pred. No. 3.3e-08;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 POEIPQQTWEEDQKAREAFRR 23  
||| ||||||||||||||||  
Db 116 PHETPQQTWEEDQKAREAFRR 138

## RESULT 13

Q67852 ID Q67852 PRELIMINARY; PRT; 327 AA.  
AC Q67852;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Surface protein.  
GN PRES.  
OS Duck hepatitis B virus (DHBV).  
OC Viruses; Retroid viruses; Hepadnaviridae; Avihepadnavirus.  
OX NCBI\_TaxID=12639;  
[1]  
RN SEQUENCE FROM N.A.  
RA Shi H., Cullen J.M., Newbold J.E.;  
RT "A novel isolate of duck hepatitis B virus";  
RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.  
RA EMBL; M95589; AAA45749.1; -;  
DR InterPro; IPR000349; Hepvir\_surfa.  
DR Pfam; PF00695; VMSA; 1.  
SQ SEQUENCE 327 AA; 36356 MW; 6875E959746DADEB CRC64;

Query Match 60.2%; Score 77; DB 12; Length 327;  
Best Local Similarity 65.2%; Pred. No. 0.011;  
Matches 15; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 POEIPQQTWEEDQKAREAFRR 23  
||| ||||||||||||  
Db 80 PPAAPVINTPEEDAKAREYFRR 102

## RESULT 14

Q9DD24 ID Q9DD24 PRELIMINARY; PRT; 104 AA.  
AC Q9DD24;  
DT 01-JUN-2001 (TREMBlrel. 17, Created)  
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE WW domain binding protein 5.  
GN WBP5.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
[1]  
RN SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=KIDNEY;  
RX MEDLINE=21085660; PubMed=11217851;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,



RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
RA Schriml L.M., Staabli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
RA Hayashizaki Y.,  
RT \*Functional annotation of a full-length mouse cDNA collection.\*;  
RL Nature 409:685-690(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE-BREAST TUMOR;  
RA Strausberg R.;  
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AK002214; BAB21940.1; -  
DR EMBL; BC007478; AAH07478.1; -  
DR MGD; MGI:109567; Wbp5.  
SQ SEQUENCE 104 AA; 12664 MW; 24512AC2DF5C1677 CRC64;  
Query Match 44.5%; Score 57; DB 11; Length 104;  
Best Local Similarity 45.5%; Pred. No. 1.7;  
Matches 10; Conservative 5; Mismatches 7; Indels 0; Gaps 0;  
Qy 1 PQEIPQOWTPEEDOKAREAFR 22  
Db 25 PEEKPEGQPEEPEEKSEETFR 46  
RESULT 15  
Q93E30  
ID Q93E30 PRELIMINARY; PRT; 311 AA.  
AC Q93E30;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE YdaO.  
GN YDAO.  
OS Salmonella enterica subsp. enterica serovar Enteritidis.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Salmonella.  
OX NCBI\_TaxID=149539;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CAHFS-285;  
RX MEDLINE=21536153; PubMed=11679316;  
RA Agron P.G., Walker R.L., Kinde H., Sawyer S.J., Hayes D.C.,  
RA Wollard J., Andersen G.L.;  
RT \*Identification by Subtractive Hybridization of Sequences Specific for  
RT Salmonella enterica Serovar Enteritidis.\*;  
RL Appl. Environ. Microbiol. 67:4984-4991(2001).  
DR EMBL; AF370716; AAL27304.1; -  
DR InterPro; IPR000541; UPF0021.  
DR Pfam; PF01171; UPF0021; 1.  
SQ SEQUENCE 311 AA; 35343 MW; 629191BF69073637 CRC64;  
Query Match 43.8%; Score 56; DB 2; Length 311;  
Best Local Similarity 60.9%; Pred. No. 7.4;  
Matches 14; Conservative 1; Mismatches 6; Indels 2; Gaps 2;  
Qy 2 QEIP-QPQ-WTPEEDOKAREAFR 22  
Db 282 EEIPLQAGWQPEEDDTALEALR 304

This Page Blank (uspto)

Copyright (c) 1993 - 2003 Compuquen Ltd.

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	871	100.0	161	2	US-08-683-262B-34	Sequence 34, Appl
2	871	100.0	161	4	US-09-361-707-34	Sequence 34, Appl
3	263	30.2	48	2	US-08-683-262B-68	Sequence 68, Appl
4	263	30.2	48	4	US-09-361-707-68	Sequence 68, Appl
5	93.5	10.7	906	2	US-08-609-230A-9	Sequence 9, Appl
6	93.5	10.7	907	3	US-08-990-140-4	Sequence 4, Appl
7	93.5	10.7	907	4	US-09-546-238-4	Sequence 4, Appl
8	93.5	10.7	940	4	US-08-810-712-7	Sequence 7, Appl
9	88.5	10.2	1194	4	US-08-538-526-1	Sequence 1, Appl
10	88	10.1	16	2	US-08-683-262B-18	Sequence 18, Appl
11	88	10.1	16	4	US-09-361-707-18	Sequence 18, Appl
12	84	9.6	417	4	US-08-887-534A-38	Sequence 38, Appl
13	84	9.6	417	4	US-08-887-534A-40	Sequence 40, Appl
14	84	9.6	417	4	US-08-887-534A-42	Sequence 42, Appl
15	83.5	9.6	605	2	US-08-687-956A-1	Sequence 1, Appl
16	82.5	9.5	1257	1	US-08-340-428B-49	Sequence 49, Appl
17	82	9.4	214	1	US-08-217-327-4	Sequence 4, Appl
18	82	9.4	1158	4	US-09-060-482-2	Sequence 2, Appl
19	82	9.4	1664	1	US-09-599-652-2	Sequence 2, Appl
20	82	9.4	1664	2	US-08-642-846-2	Sequence 2, Appl
21	82	9.4	1664	4	US-09-264-604-2	Sequence 2, Appl
22	81.5	9.4	1199	1	US-08-041-538-2	Sequence 2, Appl
23	81.5	9.4	1199	1	US-08-463-642-2	Sequence 2, Appl
24	81.5	9.4	1199	1	US-08-455-602-2	Sequence 2, Appl
25	81.5	9.4	1199	2	US-08-463-157-2	Sequence 2, Appl
26	81.5	9.4	1199	5	PCT-US91-09422-2	Sequence 2, Appl
27	81.5	9.4	1219	2	US-08-687-289A-6	Sequence 6, Appl

1 MQQHPAKSMQVRRYEGGEILLNLQAGRMTPKGTLTWSGKFPPTLDHVLDRHVTOMEINTLQ 60  
 |||||  
 1 MQQHPAKSMQVRRYEGGEILLNLQAGRMTPKGTLTWSGKFPPTLDHVLDRHVTOMEINTLQ 60  
 |||||  
 61 NQGWAPAGARRVGLSNPTPOEIQPQWTEEDQKAREFRYQERPEPTTTPPSPP 120  
 |||||

```

Db      61  NOGAWPAGARRVGLSNPTQEI PQQWTPPEEDQKAREAFRRYQERPPETTTTIPSSPP 120

QY      121  QWKLPQGGDDPLLGNQSLLETHPLYQSEPAVPVTKTPLPKKK 161
          |||||||
Db      121  QWKLPQGGDDPLLGNQSLLETHPLYQSEPAVPVTKTPLPKKK 161

RESULT 2
US-09-361-707-34
; Sequence 34, Application US/09361707
; Patent No. 6258937
; GENERAL INFORMATION:
; APPLICANT: Tong, Shuping
;           Li, Jisu
;           Wands, Jack R.
; TITLE OF INVENTION: HEPA DNA VIRUS RECEPTOR
; NUMBER OF SEQUENCES: 117
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/361,707
; FILING DATE: 27-Jul-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/683,262
; FILING DATE: 18-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Creason, Gary L.
; REGISTRATION NUMBER: 34,310
; REFERENCE/DOCKET NUMBER: 00786/287003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 161 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; SEQUENCE DESCRIPTION: SEQ ID NO: 34:
US-09-361-707-34
Query Match          100.0%; Score 871; DB 4; Length 161;
Best Local Similarity 100.0%; Pred. No. 2.8e-85;
Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY      1  MGQHPAKSMQVRRTEGGEILLNOLAGRMIPKGLTWSGKFTLHDVLDHVQMTMEINTLQ 60
Db      1  MGQHPAKSMQVRRTEGGEILLNOLAGRMIPKGLTWSGKFTLHDVLDHVQMTMEINTLQ 60

QY      61  NOGAWPAGARRVGLSNPTQEI PQQWTPPEEDQKAREAFRRYQERPPETTTTIPSSPP 120
          |||||||
Db      61  NOGAWPAGARRVGLSNPTQEI PQQWTPPEEDQKAREAFRRYQERPPETTTTIPSSPP 120

QY      121  QWKLPQGGDDPLLGNQSLLETHPLYQSEPAVPVTKTPLPKKK 161
          |||||||
Db      121  QWKLPQGGDDPLLGNQSLLETHPLYQSEPAVPVTKTPLPKKK 161

RESULT 3
US-08-683-262B-68
; Sequence 68, Application US/08683262B
; Patent No. 5929220

```

**QY** 1 MGQHPAKSMDDVRRIEGGEIL-----LNQLAGRMIPKGTLTWSGKFPTILDHVLHVDTME 54  
||| : | : ||| : : | : | :  
**Dd** 397 MGRH--RSNQLFNCHGGHHMPPTOSQFGEMCGKFMKSQGLS-----QLYHNQSOG 444

```

: TITLE OF INVENTION: Human PrtI-like Subunit Protein (hPrtI) Polynucleotides
:
: TITLE OF INVENTION: (as amended)
:
: FILE REFERENCE: 1488.0700002
:
: CURRENT APPLICATION NUMBER: US/09/546.238
:
: CURRENT FILING DATE: 2000-04-10
:
: PRIOR APPLICATION NUMBER: US 60/033.151
:
: PRIOR FILING DATE: 1996-12-13
:
: NUMBER OF SEQ ID NOS: 13
:
: SOFTWARE: PatentIn Ver. 2.1
:
: SEQ ID NO 4
:
: LENGTH: 907

```

; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-546-238-4

Query Match  
Best Local Similarity 10.7%; Score 93.5; DB 4; Length 907;  
Matches 44; Conservative 30; Mismatches 59; Indels 41; Gaps 10;  
QY 1 MGOHPAKSMVRRIRGEIL-----LNQLAGRMIPKGTLTWSGKFPTLDHVDHVTME 54  
Db 398 MGRH--RSNQLFNHGHIMPTQSGFGMGKFKKSQGLS-----QLYHNQSQG 445  
QY 55 EINTLQNGAWPAGARRVGLSNPTTQEIPOQWTPPEE-DQKARAFRYQBERP---PE 110  
Db 446 LLSQLQGGG-----KDMPPRFSSKQQLNADETSLRPAQSFLMNKNQVPLQFP 493  
QY 111 TTTTIPSS-PPQWKLP-GDDPLLGNQSLLETH-PLYQSEPAVPVVIKTPPLKKK 161  
Db 494 ITMIPPSAQPPRTQPTPLGQTPQLG----LKNPPLIQEKPAKTSKKPPPSKEE 543

## RESULT 8

US-08-810-712-7  
; Sequence 7, Application US/08810712G  
; Patent No. 6160106  
; GENERAL INFORMATION:  
; APPLICANT: Yeda Research and Development Co. LTD  
; TITLE OF INVENTION: Tumor Suppressor Genes, Proteins Encoded Thereby and  
; FILE REFERENCE: sequence list  
; CURRENT APPLICATION NUMBER: US/08/810,712G  
; CURRENT FILING DATE: 1997-03-03  
; EARLIER APPLICATION NUMBER: PCT/US94/11598  
; EARLIER FILING DATE: 1994-10-12  
; NUMBER OF SEQ ID NOS: 31  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 7  
; LENGTH: 940  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-08-810-712-7

Query Match  
Best Local Similarity 10.7%; Score 93.5; DB 4; Length 940;  
Matches 44; Conservative 30; Mismatches 59; Indels 41; Gaps 10;  
QY 1 MGOHPAKSMVRRIRGEIL-----LNQLAGRMIPKGTLTWSGKFPTLDHVDHVTME 54  
Db 431 MGRH--RSNQLFNHGHIMPTQSGFGMGKFKKSQGLS-----QLYHNQSQG 478  
QY 55 EINTLQNGAWPAGARRVGLSNPTTQEIPOQWTPPEE-DQKARAFRYQBERP---PE 110  
Db 479 LLSQLQGGG-----KDMPPRFSSKQQLNADETSLRPAQSFLMNKNQVPLQFP 526  
QY 111 TTTTIPSS-PPQWKLP-GDDPLLGNQSLLETH-PLYQSEPAVPVVIKTPPLKKK 161  
Db 527 ITMIPPSAQPPRTQPTPLGQTPQLG----LKNPPLIQEKPAKTSKKPPPSKEE 576

## RESULT 9

US-08-538-526-1  
; Sequence 1, Application US/08538526  
; Patent No. 6303751  
; GENERAL INFORMATION:  
; APPLICANT: Burnett, J. Paul  
; APPLICANT: Mayne, Nancy G.  
; APPLICANT: Sharp, Robert L.  
; APPLICANT: Snyder, Yvonne M.  
; TITLE OF INVENTION: Human Metabotropic Glutamate Receptor  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Patent Division/DKB

; STREET: Lilly Corporate Center  
; CITY: Indianapolis  
; STATE: IN  
; COUNTRY: USA  
; ZIP: 46285  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
; COMPUTER: IBM compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WordPerfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/538,526  
; FILING DATE: October 3, 1995  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Blalock, Donna K.  
; REGISTRATION NUMBER: 38,082  
; REFERENCE/DOCKET NUMBER: X-8319B  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 317/277-1090  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1194 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-538-526-1

Query Match  
Best Local Similarity 10.2%; Score 88.5; DB 4; Length 1194;  
Matches 39; Conservative 23; Mismatches 52; Indels 49; Gaps 8;  
QY 16 GGEILLNQLAGRMIPKGTLTWSGKFPTLDHVDHVTMEINTLQNGA-----WPA 67  
Db 890 GKSYSWSEPGGQVPGQGHM-----HRLSVHVKINE---TACNQTAVIKPLTKSVQ 938  
QY 58 GAGRRVGLSNPTTQEIPOQWTPPEE-DQKARAFRYQ-----ERPPTTTPSS 118  
Db 939 GSGKSLTFSDTSTKTL----YVNEEEDAQPI--RFSPGSPSMVHRRVPSAATTPPLP 992  
QY 119 PPQWKLPQDDPLLGNQSLLETHPLYQSEPAVPVVIKTPPLKKK 161  
Db 993 P-----HLTAETPLFLAEPLKGLPPLQQQ 1020

## RESULT 10

US-08-683-262B-18  
; Sequence 18, Application US/08683262B  
; Patent No. 5929220  
; GENERAL INFORMATION:  
; APPLICANT: Shuping Tong et al.  
; TITLE OF INVENTION: HEPADNAVIRUS RECEPTOR  
; NUMBER OF SEQUENCES: 75  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/683,262B  
; FILING DATE: 18-JUL-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fraser, Janis K.  
; REGISTRATION NUMBER: 31,819  
; REFERENCE/DOCKET NUMBER: 00786/287002

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 16 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-683-262B-18

Query Match 10.1%; Score 88; DB 2; Length 16;  
Best Local Similarity 100.0%; Pred. No. 0.0028;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 87 QWTPEDQKAREAFRR 102  
Db 1 QWTPEDQKAREAFRR 16  
|||||

RESULT 11  
US-09-361-707-18  
Sequence 18, Application US/09361707  
Patent No. 6258937  
GENERAL INFORMATION:  
APPLICANT: Tong, Shuping  
Li, Jisu  
Wands, Jack R.

TITLE OF INVENTION: HEP A DNA VIRUS RECEPTOR  
NUMBER OF SEQUENCES: 117  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110-2804

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
FILING DATE: 27-Jul-1999  
PRIOR APPLICATION NUMBER: US/09/361,707  
APPLICATION NUMBER: 08/683,262  
FILING DATE: 18-JUL-1996

ATTORNEY/AGENT INFORMATION:  
NAME: Creason, Gary L.  
REGISTRATION NUMBER: 34,310  
REFERENCE/DOCKET NUMBER: 00786/287003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 16 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 18:  
US-09-361-707-18

Query Match 10.1%; Score 88; DB 4; Length 16;  
Best Local Similarity 100.0%; Pred. No. 0.0028;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 87 QWTPEDQKAREAFRR 102  
Db 1 QWTPEDQKAREAFRR 16  
|||||

RESULT 12  
US-08-887-534A-38  
Sequence 38, Application US/08887534A  
Patent No. 6455323  
GENERAL INFORMATION:  
APPLICANT: Holden, David W.  
TITLE OF INVENTION: ANTI-BACTERIAL METHODS AND MATERIALS  
NUMBER OF SEQUENCES: 106  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive/6300 Sears Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/887,534A  
FILING DATE:  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Rin-Laures, Li-Hsien  
REGISTRATION NUMBER: 33,547  
REFERENCE/DOCKET NUMBER: 28341/33996  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 474-6300  
TELEFAX: (312) 474-0448  
TELEX: (312) 474-6600  
INFORMATION FOR SEQ ID NO: 38:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 417 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-887-534A-38

Query Match 9.6%; Score 84; DB 4; Length 417;  
Best Local Similarity 28.7%; Pred. No. 0.63;  
Matches 47; Conservative 20; Mismatches 67; Indels 30; Gaps 8;

Qy 14 IEGEILLNQLAGRMIPKGTLTWSGKPTLDHVL----DHVOTMEEINTLQNGAWPAGA 69  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
Db 29 VEQGEYLLEETDKVNVVELTAESG---VLQEVKDSGDTVOVGIEIGTISE-----GA 79  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
Qy 70 GRRVGLSNPTQEIPOQWTPEDQKAREAFRRYQEEPPETTTTPSSPPQWLQ----- 125  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
Db 80 GES---SAPAPTEKTESKESVKEEKQAEPAQAEVSEEAQSEAKSRTIASPSARKLAREKG 136  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
Qy 126 -----PGDDPL--LGNQSLLETHPLYQSEPAVPVPIKPPKKK 161  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
Db 137 IDLSQVPTGDPGLGRVRKQD-VEAYEKPAKPA-PQKQKQPAOK 178  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :

RESULT 13  
US-08-887-534A-40  
Sequence 40, Application US/08887534A  
Patent No. 6455323  
GENERAL INFORMATION:  
APPLICANT: Holden, David W.  
TITLE OF INVENTION: ANTI-BACTERIAL METHODS AND MATERIALS  
NUMBER OF SEQUENCES: 106  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive/6300 Sears Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
FILING DATE: US/08/887,534A  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Rin-Laures, Li-Hsien  
REGISTRATION NUMBER: 33,547  
REFERENCE/DOCKET NUMBER: 28341/33996  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 474-6300  
TELEFAX: (312) 474-0448  
INFORMATION FOR SEQ ID NO: 40:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 417 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-887-534A-40

Query Match 9.6%; Score 84; DB 4; Length 417;  
Best Local Similarity 28.7%; Pred. No. 0.63;  
Matches 47; Conservative 20; Mismatches 67; Indels 30; Gaps 8;  
QY 14 IEGGEILLNQLAGRMIPKGTLTWSGKFTLDHVL----DHVQTMBEINTLQNGAWPAGA 69  
DB 29 VEQGEVLELETDKVNVELTAESG---VLQEVLDKSGDTVQVGEIIGTISE-----GA 79  
QY 70 GRRVGLSNPTPOEIPQWPTPEEDOKAREAFRRYOERPETTTIPSPPOWKLO---- 125  
DB 80 GES---SAPAPTEKTESKVEKQAEPAQAEVSEEAQSEAKSRTIASPSARKLAREKG 136  
QY 126 -----PGDDPL--LGNQSLLETHPLYQSEPAVPVVIKTPPLKKK 161  
DB 137 IDLSQVPTGDLGRVRKQD-VEAYEKPAKPA-PQOKQOPQAK 178

RESULT 14  
US-08-887-534A-42  
Sequence 42, Application US/08887534A  
Patent No. 6455323  
GENERAL INFORMATION:  
APPLICANT: Holden, David W.  
TITLE OF INVENTION: ANTI-BACTERIAL METHODS AND MATERIALS  
NUMBER OF SEQUENCES: 106  
CORRESPONDENCE ADDRESS:  
ADDRESS: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive/6300 Sears Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
FILING DATE: US/08/887,534A  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Rin-Laures, Li-Hsien  
REGISTRATION NUMBER: 33,547  
REFERENCE/DOCKET NUMBER: 28341/33996  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 474-6300  
TELEFAX: (312) 474-0448

TELEX: (312) 474-6600  
INFORMATION FOR SEQ ID NO: 42:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 417 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-887-534A-42

Query Match 9.6%; Score 84; DB 4; Length 417;  
Best Local Similarity 28.7%; Pred. No. 0.63;  
Matches 47; Conservative 20; Mismatches 67; Indels 30; Gaps 8;  
QY 14 IEGGEILLNQLAGRMIPKGTLTWSGKFTLDHVL----DHVQTMBEINTLQNGAWPAGA 69  
DB 29 VEQGEVLELETDKVNVELTAESG---VLQEVLDKSGDTVQVGEIIGTISE-----GA 79  
QY 70 GRRVGLSNPTPOEIPQWPTPEEDOKAREAFRRYOERPETTTIPSPPOWKLO---- 125  
DB 80 GES---SAPAPTEKTESKVEKQAEPAQAEVSEEAQSEAKSRTIASPSARKLAREKG 136  
QY 126 -----PGDDPL--LGNQSLLETHPLYQSEPAVPVVIKTPPLKKK 161  
DB 137 IDLSQVPTGDLGRVRKQD-VEAYEKPAKPA-PQOKQOPQAK 178

RESULT 15  
US-08-687-956A-1  
Sequence 1, Application US/08687956A  
Patent No. 5861157  
GENERAL INFORMATION:  
APPLICANT: BURNIE, JAMES P  
APPLICANT: MATTHEWS, RUTH C  
TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF  
TITLE OF INVENTION: INFECTIONS DUE TO STREPTOCOCCI AND ENTEROCOCCI  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESS: PILLSBURY, MADISON & SUTRO, LLP  
STREET: 1100 NEW YORK AVENUE, N.W., EAST TOWER, 9TH  
CITY: WASHINGTON  
STATE: DC  
COUNTRY: USA  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
FILING DATE: 29-JUL-1996  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9401689.6  
FILING DATE: 28-JAN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: KOKULIS, PAUL N  
REGISTRATION NUMBER: 16,773  
REFERENCE/DOCKET NUMBER: 50885/222892  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202/861-3000  
TELEFAX: 202/822-0944  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 605 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: internal  
ORIGINAL SOURCE:



Search completed: January 2, 2003, 13:47:31  
Job time : 23.4234 secs

This Page Blank (uspto)

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2003, 13:45:43 ; Search time 13.6331 Seconds  
(without alignments)  
223.808 Million cell updates/sec

Title: US-09-818-066-34  
Perfect score: 871  
Sequence: 1 MGHPAKSMVRRIGEGEIL.....PLYQSEFAPVPIKTPPLKKK 161

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 117078 seqs, 18951520 residues

Total number of hits satisfying chosen parameters: 117078

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/PCCTUS\_PUBCOMB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	871	100.0	161	10	US-09-818-066-34
2	263	30.2	48	10	US-09-818-066-68
3	96	11.0	559	10	US-09-877-633-12
4	93.5	10.7	907	10	US-09-954-043-4
5	89	10.2	1337	10	US-09-803-126-1
6	88.5	10.2	586	9	US-09-764-868-624
7	88	10.1	16	10	US-09-818-066-18
8	87.5	10.0	407	10	US-09-925-301-1057
9	86	9.9	407	9	US-09-738-626-5433
10	85.5	9.8	336	10	US-09-745-763-17
11	85.5	9.8	336	10	US-09-799-777-24
12	83	9.5	442	9	US-09-738-626-5799
13	83	9.5	547	10	US-09-779-307-12
14	83	9.5	547	10	US-09-779-307-13
15	81.5	9.4	2665	10	US-09-864-761-34248
16	80.5	9.2	1596	9	US-09-902-432-4
17	80	9.2	1329	10	US-09-815-242-10112
18	79.5	9.1	344	10	US-09-771-161A-173
19	79.5	9.1	667	10	US-09-896-852-55

20	79.5	9.1	712	10	US-09-773-753-5
21	79.5	9.1	712	10	US-09-771-161A-262
22	79.5	9.1	712	10	US-09-771-161A-263
23	79.5	9.1	712	10	US-09-771-161A-264
24	79	9.1	223	10	US-09-864-761-33417
25	78.5	9.0	758	10	US-09-801-368-224
26	78.5	9.0	780	10	US-09-770-689A-5
27	78.5	9.0	1346	9	US-09-902-432-2
28	78	9.0	424	10	US-09-870-962-7
29	78	9.0	1814	10	US-09-920-552-103
30	77.5	8.9	340	9	US-09-789-054A-6
31	77.5	8.9	396	9	US-10-155-613-2
32	77.5	8.9	445	10	US-09-965-703-19
33	77.5	8.9	549	10	US-09-965-703-18
34	77.5	8.9	550	9	US-09-853-450-18
35	77.5	8.9	583	10	US-09-965-703-17
36	77.5	8.9	625	10	US-09-965-703-16
37	77.5	8.9	746	9	US-09-042-488B-5
38	77.5	8.9	746	9	US-09-042-488B-7
39	77.5	8.9	819	10	US-09-825-144-14
40	77.5	8.9	953	10	US-09-888-615-66
41	77.5	8.9	1041	9	US-09-042-488B-9
42	77	8.8	296	9	US-09-789-054A-10
43	77	8.8	521	10	US-09-749-728B-19
44	77	8.8	1056	9	US-10-161-510-10
45	76.5	8.8	1360	10	US-09-871-916-2

ALIGNMENTS

RESULT 1  
US-09-818-066-34  
Sequence 34, Application US/09818066  
Patent No. US20020032307A1  
GENERAL INFORMATION:  
APPLICANT: Shuping Tong et al.  
TITLE OF INVENTION: HEPADNAVIRUS RECEPTOR  
NUMBER OF SEQUENCES: 75  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/818,066  
FILING DATE: 27-Mar-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/683,262  
FILING DATE: 18-JUL-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Fraser, Janis K.  
REGISTRATION NUMBER: 31,819  
REFERENCE/DOCKET NUMBER: 00786/287002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 161 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 34:

Sequence 5, Appli  
Sequence 262, App  
Sequence 263, App  
Sequence 264, App  
Sequence 33417, A  
Sequence 224, App  
Sequence 5, Appli  
Sequence 7, Appli  
Sequence 103, App  
Sequence 6, Appli  
Sequence 2, Appli  
Sequence 19, Appli  
Sequence 18, Appli  
Sequence 17, Appli  
Sequence 16, Appli  
Sequence 5, Appli  
Sequence 7, Appli  
Sequence 14, Appli  
Sequence 66, Appli  
Sequence 9, Appli  
Sequence 10, Appli  
Sequence 19, Appli  
Sequence 2, Appli

US-09-818-066-34

Query Match 100.0%; Score 871; DB 10; Length 161;  
 Best Local Similarity 100.0%; Pred. No. 7e-73;  
 Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MCQHAKSMVRRIGEGEILLNQLAGRMIPKGTLTWSGKFTPLDHLVHVTMEINTLIQ 60  
 |||||  
 DB 1 MCQHAKSMVRRIGEGEILLNQLAGRMIPKGTLTWSGKFTPLDHLVHVTMEINTLIQ 60  
 |||||

QY 61 NOGAMPAGAGRRVGLSNPTPQEIPOQWTPEDQKAREAFRRYQERPPETTTIPSSPP 120  
 |||||  
 DB 61 NOGAMPAGAGRRVGLSNPTPQEIPOQWTPEDQKAREAFRRYQERPPETTTIPSSPP 120  
 |||||

QY 121 QWKLPQGGDDPLLGNSLLETHLYQSEPAVPVTKPPLKKK 161  
 |||||  
 DB 121 QWKLPQGGDDPLLGNSLLETHLYQSEPAVPVTKPPLKKK 161  
 |||||

RESULT 2

US-09-818-066-68  
 Sequence 68, Application US/09818066  
 Patent No. US20020032307A1  
 GENERAL INFORMATION:  
 APPLICANT: Shuping Tong et al.  
 TITLE OF INVENTION: HEPADNAVIRUS RECEPTOR  
 NUMBER OF SEQUENCES: 75  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Fish & Richardson P.C.  
 STREET: 225 Franklin Street  
 CITY: Boston  
 STATE: Massachusetts  
 COUNTRY: U.S.A.  
 ZIP: 02110-2804  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/818,066  
 FILING DATE: 27-Mar-2001  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/683,262  
 FILING DATE: 18-JUL-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Fraser, Janis K.  
 REGISTRATION NUMBER: 31,819  
 REFERENCE/DOCKET NUMBER: 00786/287002  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 542-5070  
 TELEFAX: (617) 542-8906  
 TELEX: 200154  
 INFORMATION FOR SEQ ID NO: 68:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 48 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 SEQUENCE DESCRIPTION: SEQ ID NO: 68:  
 US-09-818-066-68

Query Match 30.2%; Score 263; DB 10; Length 48;  
 Best Local Similarity 100.0%; Pred. No. 6.2e-18;  
 Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 AGRVGLSNPTPQEIPOQWTPEDQKAREAFRRYQERPPETTTIPP 116  
 |||||  
 DB 1 AGRVGLSNPTPQEIPOQWTPEDQKAREAFRRYQERPPETTTIPP 48  
 |||||

RESULT 3

US-09-877-633-12

Sequence 12, Application US/09877633  
 Patent No. US20020102569A1  
 GENERAL INFORMATION:  
 APPLICANT: Preeti Lal  
 APPLICANT: Jennifer Hillman  
 TITLE OF INVENTION: DIAGNOSTIC MARKER FOR CANCERS  
 FILE REFERENCE: PC-0040 CIP  
 CURRENT APPLICATION NUMBER: US/09/877,633  
 CURRENT FILING DATE: 2001-06-08  
 NUMBER OF SEQ ID NOS: 14  
 SOFTWARE: PERL Program  
 SEQ ID NO 12  
 LENGTH: 559  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: misc\_feature  
 OTHER INFORMATION: Incyte ID No. US20020102569A1 g12711367  
 US-09-877-633-12

Query Match 11.0%; Score 96; DB 10; Length 559;  
 Best Local Similarity 22.4%; Pred. No. 0.3;  
 Matches 37; Conservative 21; Mismatches 47; Indels 50; Gaps 8;

QY 5 PAKSMDVRRIGEGEILLNQLAGRMIPKGTLTWSGKFTPLDHLVHVTMEINTLONOGA 64  
 |||||  
 DB 241 PAKPQKMKTKSGPV-----MGGGLPP-----PPIKHND-----IGTWDNKG 279

QY 65 WPAGAGRRVGLSNPTPQEIPOQWTPEDQKARE-----AFRRYQE-ERPPETTTIP 115  
 |||||  
 DB 280 VPXA-----PVPQAPSPQAAPQQAQPLPAQPPALAAQPOYQSPQPPQTRWVA 330

QY 116 P-----SSPPQWKLPQGGDDPLLGNSLLETHLYQ 145  
 |||||  
 DB 331 PNRNAAFQSGGAGSDSNSPQNVQPNAP-----SVESHPVLE 369

RESULT 4

US-09-954-043-4  
 Sequence 4, Application US/09954043  
 Patent No. US20020052024A1  
 GENERAL INFORMATION:  
 APPLICANT: Olsen, Henrik S.  
 APPLICANT: Ruben, Steven M.  
 APPLICANT: Sonenberg, Nahum  
 APPLICANT: Imataka, Hiroaki  
 APPLICANT: Methot, Nathalie  
 APPLICANT: Rom, Eran  
 TITLE OF INVENTION: Human Ptrl-like Subunit Protein (hPrtl) and  
 FILE OF INVENTION: Human eIF4G-like Protein (p97) Genes  
 FILE REFERENCE: 1488.0700003  
 CURRENT APPLICATION NUMBER: US/09/954,043  
 CURRENT FILING DATE: 2001-09-18  
 PRIOR APPLICATION NUMBER: US 09/546,238  
 PRIOR FILING DATE: 2000-04-10  
 PRIOR APPLICATION NUMBER: US 08/990,140  
 PRIOR FILING DATE: 1997-12-12  
 PRIOR APPLICATION NUMBER: US 60/033,151  
 PRIOR FILING DATE: 1996-12-13  
 NUMBER OF SEQ ID NOS: 13  
 SOFTWARE: Patent In Ver. 2.1  
 SEQ ID NO 4  
 LENGTH: 907  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-09-954-043-4

Query Match 10.7%; Score 93.5; DB 10; Length 907;  
 Best Local Similarity 25.3%; Pred. No. 0.96;  
 Matches 44; Conservative 30; Mismatches 59; Indels 41; Gaps 10;

QY 1 MCQHAKSMVRRIGEGEILLNQLAGRMIPKGTLTWSGKFTPLDHLVHVTMEINTLIQ 54  
 |||||  
 DB 1 MCQHAKSMVRRIGEGEILLNQLAGRMIPKGTLTWSGKFTPLDHLVHVTMEINTLIQ 54  
 |||||

Db 398 MGRH--RSNQLFNHGHHMPTQSGCEMGKFKMSOGLS-----QLVHNSQG 445  
Qy 55 EINTLQNGAWPAGAGRRVGLSNFTQEIPOQWTPPE-DQKAREAFRRYQEEPR---PE 110  
Db 446 LLSQLQSGS-----KMPFRFSKKGQLNADEISLRPAQSFLLMKNQVPKLPQ 493  
Qy 111 TTTTPSS-PPQWKLP-GDDPLNGSLLETH-PLYQSEPAVPVVIKTPPLKKK 161  
Db 494 ITMPPSAQPPRTQPLGQTPQLG-----LKNPPLIQEKPAKTSKKPPPSKEE 543

RESULT 5

US-09-803-126-1  
; Sequence 1, Application US/09803126  
; Patent No. US20020099190A1  
; GENERAL INFORMATION:  
; APPLICANT: Brooks, Alan R.  
; APPLICANT: Deng, Gary G.  
; APPLICANT: Rubanyi, Gabor M.  
; TITLE OF INVENTION: Estrogen-Regulated Unconventional Myosin-Related  
; FILE REFERENCE: 015303-000310US  
; CURRENT APPLICATION NUMBER: US/09/803,126  
; CURRENT FILING DATE: 2001-03-09  
; PRIOR APPLICATION NUMBER: US 60/188,488  
; PRIOR FILING DATE: 2000-03-10  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 1337  
; TYPE: PRT  
; ORGANISM: Mus sp.  
; FEATURE:  
; OTHER INFORMATION: mouse myosin related protein (MRP)  
US-09-803-126-1

Query Match 10.2%; Score 89; DB 10; Length 1337;  
Best Local Similarity 21.0%; Pred. No. 4.2;  
Matches 37; Conservative 18; Mismatches 35; Indels 86; Gaps 7;  
Qy 66 PAGAGRRVGLSNFTQEIPOQWTPPEEDQKAREAFRRYQEE-----RPP----- 109  
Db 762 PSG-----VGLREDTPEAESFPQPKSFQQRDIFQKMGQDPIRVKVPKAKVQIQEE 817  
Qy 110 -----ETTTI-PPSSPPQWKLP-----GDDPLLGNQ 135  
Db 818 MEETEEDETAELSPPPPPVVKPLKASRPKRAVEDEAPAEQEEVPTQGEDPPVHSS 877  
Qy 136 SLLETHP-----LYQSEP-----AVPVIKTPPLK 159  
Db 878 NSAPQHPKPSRVPVQSSNSAPRPPQPSREIRNIIRMTQSRGPVAVPVQPTREIK 933

RESULT 6

US-09-764-868-624  
; Sequence 624, Application US/09764868  
; Patent No. US20020168711A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PFT32  
; CURRENT APPLICATION NUMBER: US/09/764,868  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - refer to PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1510  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 624  
; LENGTH: 586  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-764-868-624

Query Match 10.2%; Score 88.5; DB 9; Length 586;

Best Local Similarity 29.3%; Pred. No. 1.6;  
Matches 27; Conservative 11; Mismatches 45; Indels 9; Gaps 3;  
Qy 30 PKGTLTWSGKE--PTLDHV-----LDHVQTMEEINTLQNGAWPAGAGRRVGLSNFTQEI 83  
Db 252 PQAAIWSREINGPRHLSSPSSLESSLSTIQVDRKNGSLPCGAKELDIMSNTPPRP 311  
Qy 84 PQQWTPPEEDQKAREAFRRYQEEPPETTTIP 115  
Db 312 KPPSHLSERRQ---EENSTHSGSKKPECTLVP 340

RESULT 7

US-09-818-066-18  
; Sequence 18, Application US/09818066  
; Patent No. US20020032307A1  
; GENERAL INFORMATION:  
; APPLICANT: Shuping Tong et al.  
; TITLE OF INVENTION: HEPADNAVIRUS RECEPTOR  
; NUMBER OF SEQUENCES: 75  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/818,066  
; FILING DATE: 27-Mar-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/683,262  
; FILING DATE: 18-JUL-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fraser, Janis K.  
; REGISTRATION NUMBER: 31,819  
; REFERENCE/DOCKET NUMBER: 00786/287002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 542-5070  
; TELEFAX: (617) 542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 16 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 18:  
US-09-818-066-18

Query Match 10.1%; Score 88; DB 10; Length 16;  
Best Local Similarity 100.0%; Pred. No. 0.015;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 87 QWTPEEDQKAREAFRR 102  
Db 1 QWTPEEDQKAREAFRR 16

RESULT 8

US-09-925-301-1057  
; Sequence 1057, Application US/09925301  
; Patent No. US20020052308A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA106

```

Db 288 QPACQOQAPLPAGPPLAAGQVAPQAPLNQPTPAV 323
RESULT 10
US-09-745-763-17
; Sequence 17, Application US/09745763
; Patent No. US20020065394A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; McCoy, John M.
; Lavallee, Edward R.
; Collins-Racie, Lisa A.
; Evans, Cheryl
; Merberg, David
; Treacy, Maurice
; Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND
; ENCODING THEM
; NUMBER OF SEQUENCES: 219
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Ver
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/745,763
; FILING DATE: 18-Jun-2000
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Sprunger, Suzanne A.
; REGISTRATION NUMBER: 41,323
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8284
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 336 amino acids
; TYPE: amino acid
; STRANDEDNESS: <unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-09-745-763-17
Query Match 9.8%; Score 85.5; DB
Best Local Similarity 21.8%; Pred. No.1.4;
Matches 34; Conservative 19; Mismatches
QY 5 PAKSMVRRIRGGELLNQLAGRMTPKGLTWSGKFPILD
Db 222 PVETLSRPTVSGG---LTALPGTSPPLCVGLGSDPAPS---
QY 65 WPAGAGRRVGLSNPTQBPQPQWTPEDQKAR----EAF
Db 257 -----PSILPVPQDSPQPLPAPEEEALTTDFELL
QY 121 QWKLPQGDPLLGNSLLETHPLXGSEPAVPVTKPT 156
Db 309 ----KPPDAPPLGP-----DIHSLVQSDQEAQAVASP 336
RESULT 11
US-09-799-777-24
; Sequence 24, Application US/09799777
; Patent No. US20020091244A1

```

```

? APPLICANT: HAYASHI, MIKIRO
? APPLICANT: OCHIAI, KEIKO
? APPLICANT: YOKOI, HARUHIKO
? APPLICANT: TATEISHI, NAKO
? APPLICANT: SENO, AKIHIO
? APPLICANT: IKEDA, MASAO
? APPLICANT: OZAKI, AKIO
? TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
? FILE REFERENCE: 249-125
? CURRENT APPLICATION NUMBER: US/09/738,626
? CURRENT FILING DATE: 2000-12-18
? PRIOR APPLICATION NUMBER: JP 99/377484
? PRIOR FILING DATE: 1999-12-16
? PRIOR APPLICATION NUMBER: JP 00/159162
? PRIOR FILING DATE: 2000-04-07
? PRIOR APPLICATION NUMBER: JP 00/280988
? PRIOR FILING DATE: 2000-08-03
? NUMBER OF SEQ ID NOS: 7059
? SOFTWARE: Patentin ver. 3.0
? SEQ ID NO 5799
? LENGTH: 442
? TYPE: PRT
? ORGANISM: Corynebacterium glutamicum
US-09-738-626-5799

      Query Match          9.5%; Score 83; DB 9; Length 442;
Best Local Similarity    23.5%; Pred. No. 3.4;
Matches 42; Conservative 27; Mismatches 62; Indels 48; Gaps

QY   2  GHPAKSMVRRREGIEILLNOLAGRMIPK--GTLTWSGKFPTLDHVDHVMTMEINTLQ 60
     ||:||:|||||
Db    9  GTSPSKS-DIR-----RALPRGCTDWS-----VLPIQPVP--VEDVQ 43

QY   61  NCGANPA-CAGRRVGLSNPTEIQPQWTPRED-----QKAREAFRR----YQEERP 108
     |:||| | | | | | | | | | : | : | : | : | : | : | : | : | : | : |
Db    44  NRGAEAALDYGEKFDHIRPASVRVPAEVIAAAENTLDPLVRESIESIRRVKVAEQKP 103

QY   109 PE-TTTPSSPPQWKLO-----PGDDPLLGNOSLETHPLYQSSEPAVPVIKTTP 157
     ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    104 SEHTIELSGCIVTERFMIDRGLVYPGNAVYSSVMINTVPAQEGVNSLVVASPP 162

RESULT 13
US-09-779-307-12
? Sequence 12, Application US/09779307
? Patent No. US20020137675A1
? GENERAL INFORMATION:
? APPLICANT: Taupier Jr., Raymond
? APPLICANT: Majumder, Kumud
? APPLICANT: Vernet, Corine
? APPLICANT: Prayaga, SudhirDas
? TITLE OF INVENTION: Polynucleotides and Polypeptides Encoded Thereby
? FILE REFERENCE: 15966-662 US
? CURRENT APPLICATION NUMBER: US/09/779,307
? CURRENT FILING DATE: 2001-02-07
? PRIOR APPLICATION NUMBER: 60/180,880
? PRIOR FILING DATE: 2000-02-08
? PRIOR APPLICATION NUMBER: 60/181,044
? PRIOR FILING DATE: 2000-02-08
? PRIOR APPLICATION NUMBER: 60/181,656
? PRIOR FILING DATE: 2000-02-10
? PRIOR APPLICATION NUMBER: 60/182,795
? PRIOR FILING DATE: 2000-02-15
? NUMBER OF SEQ ID NOS: 43
? SOFTWARE: Patentin Ver. 2.1
? SEQ ID NO 12
? LENGTH: 547
? TYPE: PRT
? ORGANISM: Mus musculus
US-09-779-307-12

      Query Match          9.5%; Score 83; DB 10; Length 547;
Best Local Similarity    25.3%; Pred. No. 4.5;

```





Fri Jan 3 08:36:38 2003

us-09-818-066-34.rapb

Page 7

Search completed: January 2, 2003, 13:55:33  
Job time : 15.6331 secs

This Page Blank (uspio)

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2003, 13:41:48 ; Search time 24.0202 Seconds  
(without alignments)  
644.360 Million cell updates/sec

Title: US-09-818-066-34  
Perfect score: 871  
Sequence: 1 MGQHPAKSMVRRIEGGEIL.....PLTQSEPAVPVIRTPPLKKK 161

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- 1: pir1.\*
- 2: pir2.\*
- 3: pir3.\*
- 4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	871	100.0	364	1 SAVLD	large surface anti
2	793.5	91.1	365	1 SAVLWE	large surface anti
3	782.5	89.8	366	1 SAVLWD	large surface anti
4	777.5	89.3	366	1 SAVLBD	large surface anti
5	692	79.4	158	2 JCL095	Pre-S protein - du
6	411.5	47.2	335	1 SAVLHH	large surface anti
7	102	11.7	3530	2 A59266	unconventional myo
8	98.5	11.3	418	2 G87581	conjugal transfer
9	98	11.3	1172	2 T00065	hypothetical prote
10	96	11.0	482	2 A12094	hypothetical prote
11	95.5	11.0	633	2 T05005	hypothetical prote
12	94.5	10.8	328	2 AB0810	cell division prot
13	92	10.6	539	2 T47405	hypothetical prote
14	92	10.6	981	1 F0MVM	gag-abl polyprotei
15	92	10.6	1123	2 A39962	kinase-related tra
16	91.5	10.5	299	2 T09792	proline-rich prote
17	90	10.3	407	2 C70816	hypothetical prote
18	89	10.2	283	2 T16348	hypothetical prote
19	89	10.2	603	2 S15074	calpastatin - rat
20	89	10.2	654	2 T10772	calpastatin - rat
21	89	10.2	754	2 B97586	hypothetical prote
22	89	10.2	754	2 AC2807	OmpA family protel
23	87	10.0	296	2 A27319	gliadin - wheat
24	86.5	9.9	309	2 T29293	hypothetical prote
25	86.5	9.9	369	2 S20500	hydroxyproline-ric
26	86.5	9.9	564	2 S37241	legumin B - fava b
27	86.5	9.9	1952	2 T48814	hypothetical prote
28	86	9.9	285	2 A41826	probable pheromone
29	86	9.9	704	1 S33704	transforming prote

30 86 9.9 1473 2 A35186 salivary agglutini  
31 85.5 9.8 430 2 JC2301 hypothetical 47.8K  
32 85.5 9.8 938 1 Q0BE24 nuclear antigen EB  
33 85 9.8 224 2 T03238 extensin (clone Ex  
34 85 9.8 280 2 T03236 extensin precursor  
35 85 9.8 377 2 A48018 mucin 7 precursor,  
36 85 9.8 514 2 A56201 transcription fact  
37 85 9.8 1001 2 T28897 hypothethical prote  
38 84.5 9.7 359 2 T13478 hypothethical prote  
39 84.5 9.7 510 1 S43692 transcription fact  
40 84.5 9.7 805 2 S64238 kinesin-related pr  
41 84 9.6 379 2 T05441 proline-rich prote  
42 84 9.6 379 2 D85257 extensin-like prot  
43 84 9.6 417 2 B32879 dihydroliipoamide S  
44 84 9.6 443 2 T05540 hypothethical prote  
45 84 9.6 708 1 Q0BE88 UL80 protein - hum

ALIGNMENTS

RESULT 1

SAVLD

large surface antigen - duck hepatitis virus

N:Contains: major surface antigen; middle surface antigen

C:Species: duck hepatitis virus, DHBV

C>Date: 20-Sep-1984 #sequence\_revision 20-Sep-1984 #text\_change 13-Mar-1997

C:Accession: A03710; S12845

R:Mandart, E.; Kay, A.; Galibert, F.

J. Virol. 49, 782-792, 1984

A:Title: Nucleotide sequence of a cloned duck hepatitis B virus genome: comparison w

A:Reference number: A92997; MUID:84138772; PMID:6699938

A:Accession: A03710

A:Molecule type: DNA

A:Residues: 1-364 <MAN>

A:Cross-references: GB:R01834

R:Mates, F.; Tong, S.; Teubner, K.; Blum, H.E.

Nucleic Acids Res. 18, 6140, 1990

A:Title: Complete nucleotide sequence of a German duck hepatitis B virus.

A:Reference number: S12843; MUID:91045092; PMID:2235507

A:Accession: S12845

A:Status: translation not shown

A:Molecule type: DNA

A:Residues: 36-364 <MAT>

A:Cross-references: EMBL:X12798

C:Genetics:

A:Gene: pre-S1/pre-S2/S

C:Superfamily: hepatitis B virus surface antigen

C:Keywords: glycoprotein; surface antigen

F:89-364/Product: middle surface antigen (gene pre-S2/S) #status predicted <DSA>

F:198-364/Product: major surface antigen (gene S) #status predicted <MSA>

F:32,170,296/Binding site: carbohydrate (Asn) #status predicted

Query Match 100.0%; Score 871; DB 1; Length 364;

Best Local Similarity 100.0%; Pred. No. 9.1e-67;

Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGQHPAKSMVRRIEGGEILLNLAGRMIPKGTITWSGKFTLDHVLVDHVTMEINTLQ 60

Db 37 MGQHPAKSMVRRIEGGEILLNLAGRMIPKGTITWSGKFTLDHVLVDHVTMEINTLQ 96

Qy 61 NOGAWPAGAGRRVGLSNPTQEIFQPWTPEDQKAREAFRRYQERPEPTTTPPPSPP 120

Db 97 NOGAWPAGAGRRVGLSNPTQEIFQPWTPEDQKAREAFRRYQERPEPTTTPPPSPP 156

Qy 121 QWKLPQGDPLLGNSLLETHPLYSQSEPAVPVIRTPPLKKK 161

Db 157 QWKLPQGDPLLGNSLLETHPLYSQSEPAVPVIRTPPLKKK 197

RESULT 2

SAVLWE

large surface antigen - duck hepatitis virus (strain China)

N:Contains: major surface antigen; middle surface antigen  
C:Species: duck hepatitis virus, DHV  
C:date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 16-Jul-1999  
C:Accession: S12842  
R:Tong, S.; Mattes, F.; Teubner, K.; Blum, H.E.  
Nucleic Acids Res. 18, 6139, 1990  
A:title: Complete nucleotide sequence of a Chinese duck hepatitis B virus.  
A:reference\_number: S12840; MUID:91045091; PMID:2235506  
A:Accession: S12842  
A:Status: translation not shown  
A:Molecule\_type: DNA  
A:Residues: 1-365 <TON>  
A:cross-references: GB:M121953; NID:g325435; PIDN:AAA45746.1; PID:g325438  
C:Genetics:

Query Match	91.1%;	Score 793.5;	DB 1;	Length 365;
Best Local Similarity	90.7%;	Pred. No. 3.6e-60;		
Matches 147;	Conservative 8;	Mismatches 6;	Indels 1;	Gaps 1;

  

QY	1	MGQHPAKSMQDVR	IEGGEILLNOLAGRMTPK	GLTWSGKFPTLDHVDH	VQWMEINTLQ	60
Db	37	MGQHPAKSMQDVR	IEGGEILLNOLAGRMTPK	GAVTWSGKYPTIDHLL	DHVTWMEINTLQ	96
QY	61	NOGAWPAGAGRRV	GLNPTPOEIPQWTP	EEDOKAREFRYQ	QERPPTTTPPS	120
Db	97	QOGAWPAGAGRRV	GLNPTPOEIPQWTP	EEDOKAREFRYQ	QERPPTTTPPS	156
QY	121	QWLQPGDDP	LLNQSLLETHPLYQ	S-EPAPV	VIKTPPLKKK	161
Db	157	PWLQPGDDP	LLLEKSKLLETHPLYQ	NEPAPV	VIKTPPLKKK	198

RESULT 3  
SAVLWD  
large surface antigen - duck hepatitis virus (strain 831)  
N:Contains: major surface antigen; middle surface antigen  
C:Species: duck hepatitis virus, DHBV  
A:Note: host (Shanghai white duck)  
C:date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 04-Oct-1996  
C:Accession: D33746  
R:Uchida, M.; Esumi, M.; Shikata, T.  
Virology 173, 600-606, 1989  
Write: Molecular cloning and sequence analysis of duck hepatitis B virus  
Reference number: A33746; MUID:90085807; PMID:2596031

Query Match	89.8%;	Score 782.5;	DB 1;	Length 366;
Best Local Similarity	88.3%;	Pred. No. 3.2e-59;		
Matches 143;	Conservative 11;	Mismatches 7;	Indels 1;	Gaps
QY	1	MGQHPAKSMQVRRIEGGEILLNLQAGRMIPKGYLTWSGKFPPTLDHVLVDHVQTMEEINTLQ	60	
Db	37	MGQQPAKSMQVRRIEGCELLNLQAGRMIPKGYTWSGKFPPTIDHLLDHVQTMEEVNTLQ	96	
QY	61	NQCAWPAGACRRVGLSNPPTQETPQOWTPEEDQKAREAFRRYQERRPETTTTTPPSSPP	120	
Db	97	QCGAWPAGACRRGLGNLNPQPEPPQOWTPEEDQKAREAFRRYQERRPETTTTTIPTSTP	156	

```

Qy 121 QWKLPQGGDPLLGNQSLLETHPLYS-EPAPVPIKTPPLRKK 161
      |||||
Db 157 PWKLPQGGDPLLGNQSLLETHPLYNPEAPVPIKTPPLRKK 198
      |||||

RESULT 4
SAVLBD
large surface antigen - duck hepatitis virus (strain S5)
N/Contains: major surface antigen; middle surface antigen
C/Species: duck hepatitis virus, DHBV
A/Note: host (Shanghai brown duck)
C/Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text.
C/Accession: C3746

```

virology 1/3, 600-606, 1989  
 A;Title: Molecular cloning and sequence analysis of duck hepatitis B virus genomes of  
 A;Reference number: A33746; MUID:90085807; PMID:2596031  
 A;Accession: C33746  
 A;Molecule type: DNA  
 A;Residues: 1-366 <UCH>  
 A;Cross-references: GB:M32990  
 C;Genetics:  
 A;Gene: pre-S1/pre-S2/S  
 C;Superfamily: hepatitis B virus surface antigen  
 C;Keywords: glycoprotein; surface antigen  
 F;89-366/Product: middle surface antigen (gene pre-S2/S) #status predicted <DSA>  
 F;200-366/Product: major surface antigen (gene S) #status predicted <NSA>  
 F;170-298/Binding site: carbohydrate (Asn) (covalent) #status predicted

	Query Match	89.3%	Score 777.5	DB 1	Length 366
	Best Local Similarity	87.7%	Pred. No. 8.4e-59		
	Matches 142	Conservative 11	Mismatches 8	Indels 1	Gaps 1
Qy	1	MGQHPAKSMQVRRTEGGEILLNQLAGRMIPKGTTLTWGSKFPTDLQDLVDLHVQTMEEINTLQ	60		
Db	37	MGQHPAKSMQVRRTEGGEILLNQLAGRMIPKGTVTWGSKFPTDLQDLVDLHVQTMEEVNTMQ	96		
Qy	61	NOGAWPAGAGRRVLGNTPTQOEIPQPOWTPEDQKAEAFRRYQERPPETTTIIPSPSP	120		
Db	97	QOGAWPAGAGRRGLGNTPTQETPTQPOWTPEDQKAEAFRRYQERPPETTTIAPTST	156		
Qy	121	QWKLPQGGDDPLLGNQSLLETHPLYSQSEPAVPVTKTTPPLKKK	161		
Db	157	PWKLPQGGDDPLLGNKSLLETHPLYNQNPAPAVPVTKTTPPLKKK	198		

**Title:** Molecular cloning and sequence analysis of duck hepatitis B virus genomes of a Reference number: A33746; MUID:90085807; PMID:2596031

C;Species: duck hepatitis virus, DHAV  
C;Date: 27-Aug-1995 #sequence\_revision 27-Oct-1995 #text\_change 09-May-1997  
C;Accession: JCI095  
F;Ma, Z.M.; Li, B.L.; Xiong, S.D.; Wen, Y.M.

```

Query Match      79.4%; Score 692; DB 2; Length 158;
Best Local Similarity 100.0%; Pred. No. 5.9e-52;
Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGQHPAKSMQVRRLEGGELLNQLAGRMIPKGTITLWSGKFTLDHLVDLHDVOTMEEINTLQ 60
Db 30 MGQHPAKSMQVRRLEGGELLNQLAGRMIPKGTITLWSGKFTLDHLVDLHDVOTMEEINTLQ 89

```

Qy 61 NOGAWPAGARRVGLSNPTPOEIPQWTPEDOKAREAFRRYQERPEPTTTPSSPP 120  
Dy 90 NOGAWPAGARRVGLSNPTPOEIPQWTPEDOKAREAFRRYQERPEPTTTPSSPP 149  
Qy 121 QWKLOPG 127  
Dy 150 QWKLOPG 156

## RESULT 6

SAVLHH  
large surface antigen - heron hepatitis virus  
N:Contains: major surface antigen; middle surface antigen  
C:Species: heron hepatitis virus, HHV  
A:Note: host Ardea cinerea (gray heron)  
C:Date: 30-Sep-1989 #sequence\_revision 30-Sep-1989 #text\_change 16-Jul-1999  
C:Accession: B30082  
J. Sprengel, R.; Kaleta, E.F.; Will, H.  
J. Virol. 62, 3832-3839, 1998

A:Title: Isolation and characterization of a hepatitis B virus endemic in herons.  
A:Reference number: A93037; MUID:88333160; PMID:3418788

A:Accession: B30082  
A:Molecule type: DNA  
A:Residues: 1-335 <SPR>  
A:Cross-references: GB:M22056; NID:g325452; PIDN:AAA45739.1; PID:g325455

C:Genetics: C2685  
A:Gene: pre-S1/pre-S2/S  
C:Superfamily: hepatitis B virus surface antigen  
C:Keywords: glycoprotein; surface antigen  
F:130-335/Product: middle surface antigen (gene S) #status predicted <MSA>  
F:167-335/Product: major surface antigen (gene S) #status predicted <MSA>  
F:265/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 47.2%; Score 411.5; DB 1; Length 335;  
Best Local Similarity 48.5%; Pred. No. 1.1e-27;  
Matches 83; Conservative 23; Mismatches 48; Indels 17; Gaps 5;

Qy 1 MGQHPAKSMYRRREGGELLNQLAGMIP---KGTLTWSKFFPTLDHVLHVOTMEBIN 57  
Dy 1 MGHTQAKSTTTRRVBGGELLQHLAGMIPPEFGPIITAKGFTTQHVMDHDSVELR 60  
Qy 58 TLQOGAWPAGARRVGLSNPTPOEIPQWTPEDOKAREAFRRYQERPEPTTTP- 116  
Dy 61 TLQAGGHPEGTARRGLDQRPPT-PPITWTEEDKAKKEFFQYQENRKPARETAPP 119

Qy 117 -----SSPPQWKLOPGDDPLLGNSLLETHLYQSEPAVPVIRTPPLKKK 161  
Dy 120 ITELHAAEPPOWKISP-EDLLKAKALIPV-----KEPEVPIKVPKLTNK 164

## RESULT 7

A59266  
unconventional myosin-15 - human  
C:Species: Homo sapiens (man)  
C:Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 08-Sep-2000  
C:Accession: A59266  
F. Liang, Y.; Wang, A.; Belyantseva, I.A.; Anderson, D.W.; Probst, F.J.; Barber, T.D.; M  
an, T.B.; Fridell, R.A.

Genomics 61, 243-258, 1999  
A:Title: Characterization of the human and mouse unconventional myosin XV genes responsi

A:Reference number: A59266; MUID:20021762; PMID:10552926

A:Accession: A59266

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-3530 <LIA>

A:Cross-references: GB:AF144094; NID:g6224682; PIDN:AAF05903.1; PID:g6224683

C:Superfamily: myosin motor domain homology

F:1325-1887/Domain: myosin motor domain homology <MOS>

Query Match 11.7%; Score 102; DB 2; Length 3530;  
Best Local Similarity 31.1%; Pred. No. 4.3;  
Matches 33; Conservative 9; Mismatches 44; Indels 20; Gaps 5;

Qy 65 WPAGAGRRVGLSNTP-----QEIPQWTPED---OKAREAFRRYQERPPETT 112  
Dy 943 WPGAGRRGRSRPPVPENPFLQLGVPSPPTLPDPAADMTRVFLGRHHPGPGOLT 1002  
Qy 113 TTPSSPPQWKLOPGDDPLLGNSI-LETHLYQSEPAVPVIRTPP 157  
Dy 1003 KSAGTPE----KPEENTLGDPLPAETKP---PTAPPKDVTPP 1041

## RESULT 8

G87581

conjugal transfer protein Trb1 [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus

C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001

C:Accession: G87581

R. Nierman, W.C.; DeBry, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko

p.; Laub, M.T.; Feldblum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,

p.; J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of Caulobacter crescentus.

A:Reference number: A87249; MUID:21173698; PMID:11259647

A:Accession: G87581

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-418 <STO>

A:Cross-references: GB:AE005673; NID:gl3424269; PIDN:AAK24651.1; GSPDB:GN00148

C:Genetics:

A:Gene: CC2685

Query Match 11.3%; Score 98.5; DB 2; Length 418;

Best Local Similarity 25.4%; Pred. No. 0.66;

Matches 44; Conservative 15; Mismatches 49; Indels 65; Gaps 8;

Qy 34 LTW-----SGKFTPLDHLVHVOTMEINTLQN---QGAWPAGAG----- 70  
Dy 56 LAWFALSPRALQAQAKAPTSSAVDRGTSAAVROLPSDYAGGAPPLGPPPLDGLGRAV 115

Qy 71 -----RRVGL-----SNFTPOEIPQ-----QWTPEDOKAREAFRRYQER--- 107  
Dy 116 INHQRDGLVVGDSGSDTP--MPQSTDQQAADAEQWREGQARQAEAGVMVQVTRAE 173

Qy 108 -----PETTTIPSSPPQWKLOPGDD-----PLLGNSLLETHPL 143  
Dy 174 AAVAPNLLASEFATPPVPVSSPGIGEPGQDNKRAFNGAVPGGESIYTAHOL 226

## RESULT 9

T00065

hypothetical protein KIAA0442 - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 22-Jan-1999 #sequence\_revision 22-Jan-1999 #text\_change 21-Jul-2000

C:Accession: T00065

R. Ishikawa, K.; Nagase, T.; Nakajima, D.; Seki, N.; Ohira, M.; Miyajima, N.; Tanaka,

DNA Res. 4, 307-313, 1997

A:Title: Prediction of the coding sequences of unidentified human genes. VIII. 78 new

A:Reference number: Z14084; MUID:98116655; PMID:9455477

A:Accession: T00065

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1172 <LSH>

A:Cross-references: EMBL:AB007902; NID:g2662164; PIDN:BAA23714.1; PID:g2662165

A:Experimental source: brain

C:Genetics:

A:Note: KIAA0442

Query Match 11.3%; Score 98; DB 2; Length 1172;

Best Local Similarity 29.8%; Pred. No. 2.5;

Matches 29; Conservative 13; Mismatches 42; Indels 14; Gaps 3;

Qy 73 VGLSNTP---QEIPQWTPEDOKARE---AFRRYQERPPETTTPSSPPQWKLOP 126  
Dy 206 VLKDCPCQVQAQPIQPTQTEQLRAPSPDPLVORTAPPQPPPLSTQPPQGPPEAQLQP 265

QY 127 GDDPLL-----GNQLLETHPLYOSEPAVPVIKTP 156  
         |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  
Db 255 APQGVQRPRRPOSTTOLLHNLPPVOAHPSAQSLSOP 303

## RESULT 10

AI2094  
 hypothetical protein alr2312 [imported] - Nostoc sp. (strain PCC 7120)  
 C:Species: Nostoc sp.  
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
 C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
 C:Accession: AI2094  
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,  
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.  
 DNA Res. 8, 205-213, 2001  
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana  
 A:Reference number: AB1807; MUID:21595285; PMID:11759840  
 A:Accession: AI2094  
 A:Status: preliminary  
 A:Molecule type: DNA  
 Residues: 1-482 <KUR>  
 Cross-references: GB:BA000019; PIDN:BAW74011.1; PID:gl7131404; GSPDB:GN00179  
 A:Experimental source: strain PCC 7120  
 C:Genetics:  
 A:Gene: alr2312

Query Match	11.0%	Score 96;	DB 2;	Length 482;
Best Local Similarity	31.5%;	Pred. NO. 1.3;		
Matches 46;	Conservative	16;	Mismatches 52;	Indels 32;
				Gaps 10;

QY 23 QLAGRMIPKGTLTWSCKFPTLDHVLVDVQTMEETLQNQGAWPACAGRRVGLSNPTQE 82  
: : : : : : : : : : : : : : : :  
db 2 ELSMKRISKILL-----TNWLLISVCT--TTLITOTIAPVA-----QEDPAPE 48

QY	83	--IPQPQTPEEDOKAREAFRRYQ--EERPPETTTIIPSS--FPQWKLPQGDDPLGNQS	136
Dd	49	OTWTPOPE--NPNDKLRDALKRSPSETPPOSTTKAPESAEP--LSP--EEJAROOK	101

QY 137 LLETHPLY-----QSEPAVPVIKTP 156  
|:| | | :| |  
Db 102 LIEADKLYLAGOTAEAEKIVREVKTP 127

## RESULT 11

T05005  
hypothetical protein T19p19.70 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 11-Jun-1999  
Accession: T05005  
Bavan, M.; Monfort, A.; Casacuberta, E.; Puigdomenech, P.; Hoheisel, J.; Mewes, H.W.;  
Submitted to the Protein Sequence Database, April 1998  
A:Reference number: Z15394  
A:Accession: T05005  
A:Molecule type: DNA  
A:Residues: 1-633 <BEV>  
A:Cross-references: EMBL:AL022605  
A:Experimental source: cultivar Columbia; BAC clone T19p19  
C:Genetics:  
A:Map position: 4  
A:Introns: 385/1; 448/1; 498/3  
A:Note: T19p19.70

Query Match 11.0%; Score 95.5; DB 2; Length 633;  
Best Local Similarity 24.5%; Pred. No. 2;  
Matches 40; Conservative 15; Mismatches 45; Indels 6

[illegible]

QY 68 GAGRRV-----GLSNPTQEI PQWTPPEQDKAREAFRRYQERPPETTTIP 115

Db 520 NGRHLIAEFVRAAEVKKIEAPLP---POPOHOPOACTISR-----DDPTATP 565

QY 116 PSSPPQWKLPQGD DPLLGNQSLLETHPL YQSEPAVPVIKTPPL 158  
Dh 566 P--pp-----PLAKppHwERLPLpppppiAEEOfpPI 597

## RESEARCH 12

AB080112  
AB080110  
cell division protein [imported] - Salmonella enterica subsp. enterica serovar Typhi  
C/Species: Salmonella enterica subsp. enterica serovar Typhi  
A/Note: This species has also been called Salmonella typhi  
C/Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 09-Nov-2001  
C/Accession: AB080110  
R/Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church  
th, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr  
S.; Moule, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A/Authors: Parry, C.; Quall, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,  
A./Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se  
A/Reference number: AB0502; PMID:11677608

A:Morecore type: DNA  
A:Residues: 1-328 <PAR>  
A:Cross-references: GB:AL513382; PIDN:CAD07660.1; PID:gl6503647; GSPDB:GN00176  
C:Genetics:  
A:Gene: STV2664

Query Match	10.8%	Score	94.5	DB 2	Length	328	
Best Local Similarity	29.2%	Pred. No.	1.1				
Matches	28	Conservative	15	Mismatches	40	Gaps	4

**QY** 76 SNTPQEIPQWTP EE--DQKAREAFRYQE-----RPETTIPPSPPPQWKLPQ 127  
: : | | | | : : : : : : : : : : : :  
**Dd** 99 AQPRFAAPPOAPMGOFPVQDPVPAPQPQQVQSAPPVQPPQQPAPPSOAPOPVAOPA 158

QY 128 DDPILGNQSLLETHPLYQSEPAV---PVIKTPPLAK 160  
 159 PPP--SAOTFQPAEPVVEAPVVEKPKQK 192  
 Db

## RESULT 13

T47405  
hypothetical protein F23N14.70 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000  
C:Accession: T47405  
R:Blöcker, H.; Mewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.; Quetier, F.; Salano  
submitted to the Protein Sequence Database, March 2000  
A:Reference number: Z24465

A: Accession: T47405  
A: Status: preliminary  
A: Molecule type: DNA  
A: Residues: 1-539 <BLOS>  
A: Cross-references: EMBL:AL138638  
A: Experimental source: cultivar Columbia; BAC clone F23N14  
C: Geneticks:  
A: Map position: 3  
A: Note: F23N14.70

Query Match 10.6%; Score 92; DB 2; Length 539;  
Best Local Similarity 26.1%; Pred. No. 3.2;  
Matches 37: Conservative 16: Mismatches 63: Indels

**QY** 29 IPKGTLTWSGKF-PTLDHVLVDHVQTMEETINTLQNGAWPAGARRV-----GISNPT 79  
: : : : : : : : : : : : : : : :  
**Dd** 47 VPPSSTATMASFNPNT--RILDHRASSNR----NRGAFASAKRRRLAVDEPLDYDPLSNPA 100

QY 80 PQEIPQPQWTPEDQKAREAFRRYQEERPPETTTIPPSPPPQWKLPQGDDPLIGNQSLE 139  
- : - : : : : : : : : : : : : : : :  
Db 101 YOVVSTPLFA----SGIGSIRELLSSPPPTTS---SQPPSVSTPPPSAPPLVLSDSKD 152

QY 140 THPI.YOSEPAVPVTKTPI.KKK 161

DB 153 AEPAGLTNPASAP---PSPLAPK 171  
RESULT 14  
FOMVGM  
gag-abl polyprotein - Abelson murine leukemia virus  
N:Contains: amino end of core shell protein p30; core protein p15; inner coat protein p1  
C:Species: Abelson murine leukemia virus  
A:Note: host Mus sp. (mouse)  
C:Date: 14-Nov-1983 #sequence\_revision 09-Sep-1994 #text\_change 11-Jun-1999  
C:Accession: A03931; A0627; A93955  
R:Reddy, E.P.; Smith, M.J.; Srinivasan, A.  
Proc. Natl. Acad. Sci. U.S.A. 80, 3623-3627, 1983  
A:Title: Nucleotide sequence of Abelson murine leukemia virus genome: structural similar  
A:Reference number: A93955; MUID:83221648; PMID:6304726  
A:Accession: A03931  
A:Molecule type: DNA  
A:Residues: 1-981 <RED>  
A:Cross-references: GB:J02009; NID:9331887; PIDN:AAA46471.1; PID:9331888  
N:Note: the authors translated the codon GGA for residue 186 as Glu  
R:Reddy, E.P.; Smith, M.J.; Srinivasan, A.  
Proc. Natl. Acad. Sci. U.S.A. 80, 7372, 1983  
A:Reference number: A93980  
A:Contents: annotation; erratum, residues 588-746  
C:Genetics:  
A:Gene: gag-abl  
C:Superfamily: Abelson murine leukemia virus gag-abl polyprotein; protein kinase homolog  
C:Keywords: ATP; core protein; oncogene; phosphotransferase; polyprotein; transforming p  
F:1-131/Product: core protein p15 #status predicted <p15>  
F:132-215/Product: inner coat protein p12 #status predicted <p12>  
F:216-235/Region: amino end of core shell protein p30  
F:248-338/Domain: SH2 homology <SH2>  
F:361-621/Domain: protein kinase homology <KIN>  
F:369-377/Region: protein kinase ATP-binding motif  
F:392/Active site: Lys #status predicted  
Query Match 10.6%; Score 92; DB 1; Length 981;  
Best Local Similarity 29.0%; Pred. No. 6.6;  
Matches 31; Conservative 12; Mismatches 38; Indels 26; Gaps 5;  
QY 53 MEEINTLQNGAWPAGAGRRVGLSNPTQEIPOPOWTFEEDQKAREAFRRYQEEPPETT 112  
DB 788 MDAVNTDPTK-AGPPGE---GLRPVPPSPVKPQST-----AKPPGTP 826  
QY 113 TTPSSPPQWKLPQGDPLLGNQ--SLLETHPLYQSEPAVPVIRKTPP 157  
DB 827 TSPVSTP---STAPAPSLAGDQPPSSAAFIPLISTRVSLRKTQPP 870  
RESULT 15  
A39962  
kinase-related transforming protein (abl) (EC 2.7.1.-) type I - mouse  
N:Alternate names: protein-tyrosine kinase abl  
C:Species: Mus musculus (house mouse)  
C:Date: 08-Nov-1991 #sequence\_revision 08-Nov-1991 #text\_change 18-Jun-1999  
C:Accession: A39962; A24773; S00771; A00626  
R:Oppi, C.; Shore, S.K.; Reddy, E.P.  
Proc. Natl. Acad. Sci. U.S.A. 84, 8200-8204, 1987  
A:Title: Nucleotide sequence of testis-derived c-abl cDNAs: implications for testis-spec  
A:Reference number: A39962; MUID:88068561; PMID:3317402  
A:Accession: A39962  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1123 <OPP>  
A:Cross-references: GB:J02995; NID:g191566; PIDN:AAA88241.1; PID:g309084  
A:Experimental source: Clone A16  
R:Ben-Neriah, Y.; Bernards, A.; Paskind, M.; Daley, G.Q.; Baltimore, D.  
Cell 44, 577-586, 1986  
A:Title: Alternative 5' exons in c-abl mRNA.  
A:Reference number: A90877; MUID:86133550; PMID:3512096  
A:Accession: A24773  
A:Molecule type: mRNA

A:Residues: 1-118 <BEN>  
R:Bernards, A.; Paskind, M.; Baltimore, D.  
Oncogene 2, 297-304, 1988  
A:Title: Four murine c-abl mRNAs arise by usage of two transcriptional promoters and  
A:Reference number: S00771; MUID:88202920; PMID:3283651  
A:Accession: S00771  
A:Molecule type: DNA  
A:Residues: 1-26 <BER>  
A:Cross-references: EMBL:X07539; NID:g49837; PIDN:CAA30411.1; PID:g49838  
R:Wang, J.Y.J.; Ledley, F.; Goff, S.; Lee, R.; Groner, Y.; Baltimore, D.  
Cell 36, 349-356, 1984  
A:Title: The mouse c-abl locus: molecular cloning and characterization.  
A:Reference number: A00626; MUID:84108840; PMID:6319018  
A:Accession: A00626  
A:Molecule type: DNA  
A:Residues: 85-127, 'C', 129-182 <WAN>  
C:Genetics:  
A:Gene: abl  
C:Superfamily: human protein-tyrosine kinase abl; protein kinase homology; SH2 homolo  
C:Keywords: alternative splicing; ATP; autophosphorylation; nucleus; phosphoprotein;  
F:68-116/Domain: SH3 homology <SH3>  
F:127-217/Domain: SH2 homology <SH2>  
F:240-500/Domain: protein kinase homology <KIN>  
F:248-256/Region: protein kinase ATP-binding motif  
Query Match 10.6%; Score 92; DB 2; Length 1123;  
Best Local Similarity 29.0%; Pred. No. 7.8;  
Matches 31; Conservative 12; Mismatches 38; Indels 26; Gaps 5;  
QY 53 MEEINTLQNGAWPAGAGRRVGLSNPTQEIPOPOWTFEEDQKAREAFRRYQEEPPETT 112  
DB 930 MDAVNTDPTK-AGPPGE---GLRPVPPSPVKPQST-----AKPPGTP 968  
QY 113 TTPSSPPQWKLPQGDPLLGNQ--SLLETHPLYQSEPAVPVIRKTPP 157  
DB 969 TSPVSTP---STAPAPSLAGDQPPSSAAFIPLISTRVSLRKTQPP 1012  
Search completed: January 2, 2003, 13:45:33  
Job time : 27.0202 secs

This Page Blank (uspto)



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2003, 13:40:48 ; Search time 12.9839 Seconds  
(without alignments)  
514.306 Million cell updates/sec

Title: US-09-818-066-34

Perfect score: 871

Sequence: 1 MGQHPAKSMVRRIEGEIL.....PLYQSEPAVPVITPPLKKK 161

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	871	100.0	328	1 VMSA_HPBDU	P03145 duck hepati
2	793.5	91.1	365	1 VMSA_HPBDC	P30029 duck hepati
3	782.5	89.8	366	1 VMSA_HPBDB	P17195 duck hepati
4	777.5	89.3	366	1 VMSA_HPBDB	P17194 duck hepati
5	411.5	47.2	335	1 VMSA_HPBHE	P13847 heron hepat
6	102	11.7	3530	1 MY15_HUMAN	Q9ukn7 homo sapien
7	94.5	10.8	328	1 ZTPA_SALTY	P55894 salmonella
8	92	10.6	746	1 ABL_MLVAB	P00521 abelson mur
9	92	10.6	1123	1 ABL_MOUSE	P00520 mus musculu
10	89	10.2	654	1 ICAL_RAT	P27321 rattus norv
11	89	10.2	1319	1 MN1_HUMAN	Q10571 homo sapien
12	88.5	10.2	1194	1 MGRI_HUMAN	Q13255 homo sapien
13	86	9.9	704	1 MYBB_MOUSE	P48972 mus musculu
14	86	9.9	1500	1 SSP5_STRGN	P16952 streptococc
15	85.5	9.8	938	1 EBNA_EBV	P03203 Epstein-bar
16	85	9.8	514	1 MEFD_MOUSE	Q63943 mus musculu
17	85	9.8	565	1 MOT8_MOUSE	Q70324 mus musculu
18	84.5	9.7	510	1 ERM_HUMAN	P41161 homo sapien
19	84.5	9.7	805	1 YGW6_YEAST	P53086 saccharomyc
20	84	9.6	417	1 ODO2_BACSU	P16263 bacillus su
21	84	9.6	708	1 VP40_HCNVA	P16733 human cytom
22	84	9.6	1902	1 SMF1_HUMAN	Q14497 homo sapien
23	83	9.5	407	1 ARRH_LOCF1	P32122 locusta mig
24	82.5	9.5	757	1 HT16_HYDAT	P53356 hydra atten
25	82.5	9.5	1257	1 PCN_RAT	P55067 rattus norv
26	82	9.4	309	1 NO75_SOYBN	P08297 glycine max
27	82	9.4	1664	1 INT1_CANAL	P53705 candida alb
28	82	9.4	5376	1 ZAN_MOUSE	O88799 mus musculu
29	81.5	9.4	1199	1 MGRI_RAT	P23385 rattus norv
30	81	9.3	2469	1 TEGU_HSVSA	Q01056 herpesvirus
31	80.5	9.2	269	1 YA53_ARCFU	O29209 archaeoglob
32	80.5	9.2	484	1 LEGA_VICFA	P05190 vicia faba
33	80.5	9.2	1001	1 IF2_SYNT3	P72689 synechocyst

#### RESULT 1

ID	VMSA_HPBDU	STANDARD;	PRT;	328 AA.
AC	P03145;			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	01-JUL-1993 (Rel. 26, Last sequence update)			
DT	01-FEB-1994 (Rel. 28, Last annotation update)			
DE	Major surface antigen precursor.			
GN	S.			
OS	Duck hepatitis B virus (DHBV).			
OC	Viruses; Retroviridae; Hepadnaviridae; Avihepadnavirus.			
OX	NCBI_TaxID=12639;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=84138772; PubMed=6699938;			
RA	Mandart E., Kay A., Galibert F.;			
RT	"Nucleotide sequence of a cloned duck hepatitis B virus genome:			
RT	comparison with woodchuck and human hepatitis B virus sequences.";			
RL	J. Virol. 49:782-792(1984).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Isolate DHBV Fl-6;			
RX	MEDLINE=91045092; PubMed=2235507;			
RA	Mattes F., Tong S., Teubner K., Altm H.E.;			
RT	"Complete nucleotide sequence of a German duck hepatitis B virus.";			
RL	Nucleic Acids Res. 18:6140-6140(1990).			
RP	[3]			
RX	MYRISTOYLATION.			
RA	MEDLINE=91135002; PubMed=1994583;			
RA	Macrae D.R., Bruss V., Ganem D.;			
RT	"Myristylation of a duck hepatitis B virus envelope protein is			
RT	essential for infectivity but not for virus assembly.";			
RL	Virology 181:359-363(1991).			
CC	-[- PTM: MYRISTOYLATION CONTRIBUTES IMPORTANTLY TO DHBV INFECTIVITY.			
CC	IT IS MOST LIKELY REQUIRED FOR AN EARLY STEP OF THE LIFE CYCLE			
CC	INVOLVING THE ENTRY OR UNCOATING OF VIRUS PARTICLES.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; X12798; CAB57224.1; -			
DR	PIR; A03710; SIVLD.			
DR	PIR; S12845; S12845			
DR	InterPro: IPR000349; Hepvir_surfa9.			
DR	Pfam; PF00695; VMSA; 1.			
KW	Antigen; Myristate; Envelope protein; Lipoprotein.			
FT	PROPEP 1 161			
FT	CHAIN 162 328			
FT	LIPID 2 2			
FT	CARBOHYD 134 134			
FT	CARBOHYD 260 260			
FT	MAJOR SURFACE ANTIGEN.			
FT	MYRISTATE.			
FT	N-LINKED (GLCNAC. . .) (POTENTIAL).			
FT	N-LINKED (GLCNAC. . .) (POTENTIAL).			

P04726 triticum ae  
P46889 escherichia  
Q49429 mycoplasma  
Q9um66 homo sapien  
P11716 oryctolagus  
P51617 homo sapien  
O95104 homo sapien  
O94916 homo sapien  
P42694 homo sapien  
Q03825 saccharomyc  
O13415 aspergillus  
Q8zsv6 pyrobaculum

```
SQ SEQUENCE 328 AA; 36230 MW; B2D771241E407456 CRC64;

Query Match
Best Local Similarity 100.0%; Score 871; DB 1; Length 328;
Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQGHPAKSMVRRIRGGELLNQLAGRMIPKGTLTWSGKFFPTLDHVLHVQTMEEINTLIQ 60
DB 1 MQGHPAKSMVRRIRGGELLNQLAGRMIPKGTLTWSGKFFPTLDHVLHVQTMEEINTLIQ 60
QY 61 NOGAWPAGAGRRVGLSNTPTQEIPOQWTPEDQKARAFRRYQERPPETTTIPSSPP 120
DB 61 NOGAWPAGAGRRVGLSNTPTQEIPOQWTPEDQKARAFRRYQERPPETTTIPSSPP 120
QY 121 QWKLPQGGDDPLLGNSLLETHLYQSEPAVPVKTPLKKK 161
DB 121 QWKLPQGGDDPLLGNSLLETHLYQSEPAVPVKTPLKKK 161

RESULT 2
VMSA_HPBDC STANDARD; PRT; 365 AA.
AC P30029;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Major surface antigen precursor.
GN S.
OS Duck hepatitis B virus (strain China) (DHBV).
OC Viruses; Retroviral viruses; Hepadnaviridae; Avihepadnavirus.
OX NCBI_TaxID=31510;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91045091; PubMed=2235506;
RA Tong S., Maties F., Teubner K., Blum H.E.;
RT "Complete nucleotide sequence of a Chinese duck hepatitis B virus.";
RL Nucleic Acids Res. 18:6139-6139(1990).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M21953; AAA45746.1;
DR PIR; S12842; SAVLWE.
DR InterPro; IPR000349; Hepvir_surfAg.
DR Pfam; PF00695; VMSA; 1.
KW Antigen.
FT PROPEP 1 198
FT CHAIN 199 365 MAJOR SURFACE ANTIGEN.
FT CARBOHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT SEQUENCE 365 AA; 40511 MW; 67F3A4174CB/D884 CRC64;

Query Match
Best Local Similarity 91.1%; Score 793.5; DB 1; Length 365;
Matches 147; Conservative 8; Mismatches 6; Indels 1; Gaps 1;

QY 1 MQGHPAKSMVRRIRGGELLNQLAGRMIPKGTLTWSGKFFPTLDHVLHVQTMEEINTLIQ 60
DB 37 MQGHPAKSMVRRIRGGELLNQLAGRMIPKGTLTWSGKFFPTLDHVLHVQTMEEINTLIQ 96
QY 61 NOGAWPAGAGRRVGLSNTPTQEIPOQWTPEDQKARAFRRYQERPPETTTIPSSPP 120
DB 97 QGAWPAGAGRRVGLSNTPTQEIPOQWTPEDQKARAFRRYQERPPETTTIPSSPP 156
QY 121 QWKLPQGGDDPLLGNSLLETHLYQSEPAVPVKTPLKKK 161
DB 121 QWKLPQGGDDPLLGNSLLETHLYQSEPAVPVKTPLKKK 198

RESULT 3
VMSA_HPBDC STANDARD; PRT; 366 AA.
AC P17195;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 01-MAY-1991 (Rel. 18, Last annotation update)
DE Major surface antigen precursor.
GN S.
OS Duck hepatitis B virus (white Shanghai duck isolate s31) (DHBV).
OC Viruses; Retroviral viruses; Hepadnaviridae; Avihepadnavirus.
OX NCBI_TaxID=10440;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90085807; PubMed=2596031;
RA Uchida M., Esumi M., Shikata T.;
RT "Molecular cloning and sequence analysis of duck hepatitis B virus
RT genomes of a new variant isolated from Shanghai ducks.";
RL Virology 173:600-606(1989).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M32991; AAA45752.1; ALT_INIT.
DR PIR; D33746; SAVLWD.
DR InterPro; IPR000349; Hepvir_surfAg.
DR Pfam; PF00695; VMSA; 1.
KW Antigen.
FT PROPEP 1 199
FT CHAIN 200 366 MAJOR SURFACE ANTIGEN.
FT CARBOHYD 170 170 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 298 298 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT SEQUENCE 366 AA; 40858 MW; CF60E78B7B2FCD52 CRC64;

Query Match
Best Local Similarity 89.8%; Score 782.5; DB 1; Length 366;
Matches 143; Conservative 11; Mismatches 7; Indels 1; Gaps 1;

QY 1 MQGHPAKSMVRRIRGGELLNQLAGRMIPKGTLTWSGKFFPTLDHVLHVQTMEEINTLIQ 60
DB 37 MQGHPAKSMVRRIRGGELLNQLAGRMIPKGTLTWSGKFFPTLDHVLHVQTMEEINTLIQ 96
QY 61 NOGAWPAGAGRRVGLSNTPTQEIPOQWTPEDQKARAFRRYQERPPETTTIPSSPP 120
DB 97 QGAWPAGAGRRVGLSNTPTQEIPOQWTPEDQKARAFRRYQERPPETTTIPSSPP 156
QY 121 QWKLPQGGDDPLLGNSLLETHLYQSEPAVPVKTPLKKK 161
DB 157 PWKLPQGGDDPLLGNSLLETHLYQSEPAVPVKTPLKKK 198

RESULT 4
VMSA_HPBDC STANDARD; PRT; 366 AA.
AC P17194;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 01-MAY-1991 (Rel. 18, Last annotation update)
DE Major surface antigen precursor.
GN S.
OS Duck hepatitis B virus (brown Shanghai duck isolate S5) (DHBV).
OC Viruses; Retroviral viruses; Hepadnaviridae; Avihepadnavirus.
OX NCBI_TaxID=10439;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90085807; PubMed=2596031;
RA Uchida M., Esumi M., Shikata T.;
RT "Molecular cloning and sequence analysis of duck hepatitis B virus
```





QY 128 DPLILGNQLLTHPLYOSEPAV---PVIKTPPLKK 160  
 ID ABL\_MLVB STANDARD; PRT; 746 AA.  
 AC P00521;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Tyrosine-protein kinase transforming protein ABL (EC 2.7.1.112).  
 GN V-ABL.  
 OS Abelson murine leukemia virus.  
 OC Viruses; Retroviral viruses; Retroviridae; Mammalian type C retroviruses.  
 OX NCBI\_TaxID=11788;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=83221648; PubMed=6304726;  
 RA Reddy E.P., Smith M.J., Srinivasan A.;  
 RT "Nucleotide sequence of Abelson murine leukemia virus genome:  
 structural similarity of its transforming gene product to other onc  
 gene products with tyrosine-specific kinase activity.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 80:3623-3627(1983).  
 RN [2]  
 RP REVISIONS TO 588-746.  
 RA Reddy E.P., Smith M.J., Srinivasan A.;  
 RL Proc. Natl. Acad. Sci. U.S.A. 80:7372-7372(1983).  
 RN [3]  
 RP SEQUENCE OF 233-327 FROM N.A.  
 RX MEDLINE=83245023; PubMed=6191223;  
 RA Groffen J., Heisterkamp N., Reynolds F.H. Jr., Stephenson J.R.;  
 RT "Homology between phosphotyrosine acceptor site of human c-abl and  
 viral oncogene products.";  
 RL Nature 304:167-169(1983).  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine - ADP + protein  
 tyrosine phosphate.  
 CC -1- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ABL  
 POLYPEPTIDE.  
 CC -1- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES. ABL  
 SUBFAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; V01541; -; NOT ANNOTATED\_CDS.  
 CC EMBL; K00010; AAA46470.1; -;  
 CC PIR; A0627; TVAVGM.  
 CC HSP; P00519; 2ABL.  
 CC InterPro; IPR000719; Euk\_pkinase.  
 CC InterPro; IPR000980; SH2.  
 CC InterPro; IPR001245; Tyr\_pkinase.  
 CC Pfam; PF00017; SH2; 1.  
 CC Pfam; PF00069; Pkinase; 1.  
 CC PRINTS; PR00401; SH2DOMAIN.  
 CC PRINTS; PR00109; TYRKINASE.  
 CC ProDom; PD000001; Euk\_pkinase; 1.  
 CC ProDom; PD000093; SH2; 1.  
 CC SMART; SM00252; SH2; 1.  
 CC SMART; SM00219; TyrKc; 1.  
 CC PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 CC PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 CC PROSITE; PS00011; PROSITE\_TYR\_KINASE\_DOM; 1.  
 CC PROSITE; PS50001; SH2; 1.  
 CC Polyprotein; Tyrosine-protein kinase; Transferase; Oncogene;

KW SH2 domain. 13 103 SH2.  
 FT DOMAIN 128 379 PROTEIN KINASE.  
 SQ SEQUENCE 746 AA; 81872 MW; B9073FF5FE9257 CRC64;  
 Query Match 10.68; Score 92; DB 1; Length 746;  
 Best Local Similarity 29.08; Pred. No. 3.2;  
 Matches 31; Conservative 12; Mismatches 38; Indels 26; Gaps 5;  
 QY 53 MEEINTLQAGAPAGARRVGLSNPTQEIPOQPOWTEEDQKAREAFRRYOEPPPTT 112  
 ID ABL\_MOUSE STANDARD; PRT; 1123 AA.  
 AC P00520; Q61252; Q61253; Q61254; Q61255; Q61256; Q61257; Q61258;  
 AC Q61259; Q61260; Q61261; P97896;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Proto-oncogene tyrosine-protein kinase ABL1 (EC 2.7.1.112) (p150)  
 DE (C-ABL).  
 GN ABL1 OR ABL.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RX MEDLINE=86068561; PubMed=3317402;  
 RA Oppi C., Shore S.K., Reddy E.P.;  
 RT "Nucleotide sequence of testis-derived c-abl cDNAs: Implications for  
 testis-specific transcription and abl oncogene activation.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 84:8200-8204(1987).  
 RN [2]  
 RP SEQUENCE OF 1-187 FROM N.A. (ISOFORMS I TO IV).  
 RX MEDLINE=95394474; PubMed=7665185;  
 RA Chisoso S.L., Bodenteich A., Wang Y.-F., Wang Y.-P., Burian D.,  
 RA Clifton S.W., Crabtree J., Freeman A., Iyer K., Jian L., Ma Y.,  
 RA McLaury H.-J., Pan H.-O., Sarhan O.H., Toth S., Wang Z., Zhang G.,  
 RA Heisterkamp N., Groffen J., Roe B.A.;  
 RT "Sequence and analysis of the human ABL gene, the BCR gene, and  
 regions involved in the Philadelphia chromosome translocation.";  
 RL Genomics 27:67-82(1995).  
 RN [3]  
 RP SEQUENCE OF 85-182 FROM N.A.  
 RX MEDLINE=84106840; PubMed=6319018;  
 RA Wang J.Y.J., Ledley F., Goff S., Lee R., Groner Y., Baltimore D.;  
 RT "The mouse c-abl locus: molecular cloning and characterization.";  
 RL Cell 36:349-356(1984).  
 RN [4]  
 RP ALTERNATIVE SPLICING.  
 RX MEDLINE=88202920; PubMed=3283651;  
 RA Bernards A., Paskind M., Baltimore D.;  
 RT "Four murine c-abl mRNAs arise by usage of two transcriptional  
 promoters and alternative splicing.";  
 RL Oncogene 2:297-304(1988).  
 RN [5]  
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 61-121.  
 RX MEDLINE=95393198; PubMed=7664083;  
 RA Musacchio A., Saraste M., Wilmanns M.;  
 RT "High-resolution crystal structures of tyrosine kinase SH3 domains  
 complexed with proline-rich peptides.";  
 RL Nat. Struct. Biol. 1:546-551(1994).  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine - ADP + protein  
 tyrosine phosphate.



[illegible]

```

RESULT 12
MGRI_HUMAN
ID MGRI_HUMAN STANDARD; PRT; I194 AA.
Q13255; Q14757; Q14758; Q9UGT0; Q9UGS9; Q9NU10;
AC 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Metabotropic glutamate receptor 1 precursor (mGluR1).
GN GRM1 OR GPRC1A OR MGLUR1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
RN [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=96029774; PubMed=7476890;
RX Desai M.A., Burnett J.P., Mayne N.G., Schoepp D.D.;
RT "Cloning and expression of a human metabotropic glutamate receptor 1
RT alpha: enhanced coupling on co-transfection with a glutamate
RT transporter.";
RT Mol. Pharmacol. 48:648-657(1995).
RL [2]
RN SEQUENCE FROM N.A.
RP MEDLINE=97231349; PubMed=9076744;
RX Stephan D., Bon C., Holzwarth J.A., Galvan M., Pruss R.M.;
RT "Human metabotropic glutamate receptor 1: mRNA distribution,
RT chromosome localization and functional expression of two splice
RT variants.";
RT Neuropharmacology 35:1649-1660(1996).
RL [3]
RN SEQUENCE FROM N.A.
RP Bates K.;
RA Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: RECEPTOR FOR GLUTAMATE. THE ACTIVITY OF THIS RECEPTOR IS
CC MEDIATED BY A G-PROTEIN THAT ACTIVATES A PHOSPHATIDYLINOSITOL-
CC CALCIUM SECOND MESSENGER SYSTEM. MAY PARTICIPATE IN THE CENTRAL
CC ACTION OF GLUTAMATE IN THE CNS, SUCH AS LONG-TERM POTENTIATION IN
CC THE HIPPOCAMPUS AND LONG-TERM DEPRESSION IN THE CEREBELLUM.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; ALPHA (SHOWN HERE) AND BETA;
CC ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: BELONGS TO FAMILY 3 OF G-PROTEIN COUPLED RECEPTORS.

```



STRONGEST, TO MGLURS.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL: U31215; AAB87843.1; -  
EMBL: U31216; AAB87844.1; -  
EMBL: L76627; AAB05337.1; -  
EMBL: L76631; AAB05338.1; -  
EMBL: AL096867; CAB75694.1; -  
EMBL: AL035698; CAB65991.1; -  
EMBL: AL035698; CAB65992.1; -  
Genew; HGNC:4593; GRM1.  
MIM; 60473; -  
InterPro; IPR001828; ANF\_receptor.  
InterPro; IPR000337; GPCR\_Mgr.  
Pfam; PF00003; 7tm3; 1.  
Pfam; PF01094; ANF\_receptor; 1.  
PRINTS; P00248; GPCRGR.  
PROSITE; PS00979; G\_PROTEIN\_RECP\_F3\_1; 1.  
PROSITE; PS00980; G\_PROTEIN\_RECP\_F3\_2; 1.  
PROSITE; PS00981; G\_PROTEIN\_RECP\_F3\_3; 1.  
PROSITE; PS0259; G\_PROTEIN\_RECP\_F3\_4; 1.  
G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;  
Multigene family; Alternative splicing.  
SIGNAL 1 18  
CHAIN 19 1194 METABOTROPIC GLUTAMATE RECEPTOR 1.  
DOMAIN 19 592 EXTRACELLULAR (POTENTIAL).  
DOMAIN 593 615 I (POTENTIAL).  
DOMAIN 616 629 CYTOPLASMIC (POTENTIAL).  
DOMAIN 630 650 III (POTENTIAL).  
DOMAIN 651 661 EXTRACELLULAR (POTENTIAL).  
DOMAIN 662 680 III (POTENTIAL).  
DOMAIN 681 706 CYTOPLASMIC (POTENTIAL).  
DOMAIN 707 727 IV (POTENTIAL).  
DOMAIN 728 750 EXTRACELLULAR (POTENTIAL).  
DOMAIN 751 772 V (POTENTIAL).  
DOMAIN 773 785 CYTOPLASMIC (POTENTIAL).  
DOMAIN 786 808 VI (POTENTIAL).  
DOMAIN 809 814 EXTRACELLULAR (POTENTIAL).  
DOMAIN 815 840 VII (POTENTIAL).  
DOMAIN 841 1194 CYTOPLASMIC (POTENTIAL).  
DOMAIN 1014 1035 GLN/PRO-RICH.  
DOMAIN 1067 1081 GLN/PRO-RICH.  
DOMAIN 1095 1130 ASP/GLU-RICH (ACIDIC).  
DOMAIN 1142 1194 SER-RICH.  
CARBOHYD 98 98 N-LINKED (GLCNAC. .) (POTENTIAL).  
CARBOHYD 223 223 N-LINKED (GLCNAC. .) (POTENTIAL).  
CARBOHYD 397 397 N-LINKED (GLCNAC. .) (POTENTIAL).  
CARBOHYD 515 515 N-LINKED (GLCNAC. .) (POTENTIAL).  
VARFPLIC 887 906 NSNGSVSWSEGGQVPRG -> KKQPEFSPSQCPSSAH  
VOL (IN ISOFORM BETA).  
VARFPLIC 907 1194 MISSING (IN ISOFORM BETA).  
CONFLICT 593 593 P -> S (IN REF. 2 AND 3).  
CONFLICT 993 993 P -> S (IN REF. 3).  
SEQUENCE 1194 AA; 132376 MW; 970E51AF40584F40 CRC64;  
Query Match 10.2%; Score 88.5; DB 1; Length 1194;  
Best Local Similarity 23.9%; Pred. No. 11;  
Matches 39; Conservative 23; Mismatches 52; Indels 49; Gaps 8;  
QY 16 GGEILLNQLAGRMIPKGTITWSKPTLDHVLVD-HVQMTSEINTLQNOCA-----WPA 67  
Db 890 GKSWSWSEGGQVPRGQHW-----HRLSVHVKTNE---TACNQFAVIKPTKSYQ 938  
QY 68 GAGRRVGLSNPTQEIPOQWTPTEEDQKAREAFRRYQ-----EERPPETTTTPSS 118  
Db 939 GSKSITFSDTSTKTL-----YVNEEEDAQPI--RSPDPSMVVHRRVPSAATTPPLP 992

QY 119 PPQWKLPQGDPLGNQSLLETHPLYQSEPAVPVITKTPPLKKK 161  
Db 993 P-----HLTAETPLFLAEPALPKGLPPPLQQQ 1020  
RESULT 13  
MYBB\_MOUSE STANDARD; PRT; 704 AA.  
AC P48972;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Myb-related protein B (B-Myb).  
GN MYBL2 OR BMYB.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92366176; PubMed=1501895;  
RA Lam E.W., Robinson C., Watson R.J.;  
RT "Characterization and cell cycle-regulated expression of mouse  
B-myb.";  
RL Oncogene 7:1885-1890(1992).  
RN [2]  
RP SEQUENCE OF 1-6 FROM N.A.  
RX MEDLINE=93327760; PubMed=8334989;  
RA Lam E.W., Watson R.J.;  
RT "An E2F-binding site mediates cell-cycle regulated repression of  
mouse B-myb transcription.";  
RL EMBO J. 12:2705-2713(1993).  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- PTM: PHOSPHORYLATED BY CYCLIN A/CDK2 DURING S-PHASE.  
CC PHOSPHORYLATION AT THR-524 IS PROBABLY INVOLVED IN TRANSCRIPTIONAL  
ACTIVITY (BY SIMILARITY).  
CC -!- SIMILARITY: CONTAINS 3 MYB-LIKE DOMAINS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL: X70472; CAA49898.1; -  
EMBL: X73028; CAA51511.1; -  
HSP; Q03237; IASJ.  
MGI:101785; Mybl2.  
InterPro; IPR001005; Myb\_DNA\_binding.  
Pfam; PF00249; myb\_DNA-binding; 3.  
SMART; SM00395; SANT; 3.  
PROSITE; PS00037; MYB\_1; 3.  
PROSITE; PS00334; MYB\_2; 3.  
PROSITE; PS00090; MYB\_3; 3.  
Transcription regulation; Nuclear protein; DNA-binding; Repeat;  
Phosphorylation.  
DNA\_BIND 26 77 MYB 1.  
DNA\_BIND 78 129 MYB 2.  
DNA\_BIND 130 180 MYB 3.  
MOD\_RES 443 443  
MOD\_RES 447 447  
MOD\_RES 490 490  
MOD\_RES 497 497  
MOD\_RES 524 524  
MOD\_RES 581 581



```

DR ProDom; PD153432; Csurface_antigen; 2.  

DR TIGRFAMS; TIGR01167; LPXTG_anchor; 1.  

DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.  

KW Cell wall; Peptidoglycan-anchor; Calcium-binding; Repeat; Signal.  

FT SIGNAL 1 38  

FT CHAIN 39 1470 AGGLUTININ RECEPTOR.  

FT PROPEP 1471 1500 REMOVED BY SORTASE (POTENTIAL).  

FT SIMILAR 164 470 TO M PROTEIN OF S.PYOGENES.  

FT DOMAIN 164 470 4 X APPROXIMATE TANDEM REPEATS, HRI.  

FT REPEAT 164 241 1.  

FT REPEAT 242 323 2.  

FT REPEAT 324 405 3.  

FT REPEAT 406 470 4.  

FT DOMAIN 771 887 3 X APPROXIMATE TANDEM REPEATS, PRI.  

FT CA_BIND 1414 1436 PRO-RICH (PR2).  

FT CA_BIND 220 235 POTENTIAL.  

FT CA_BIND 301 316 POTENTIAL.  

FT CA_BIND 931 950 POTENTIAL.  

FT CA_BIND 1300 1315 POTENTIAL.  

FT SITE 1467 1471 LPXTG SORTING SIGNAL (POTENTIAL).  

FT MOD_RES 1470 1470 AMIDE-LINKED TO CELL WALL (POTENTIAL).  

SQ SEQUENCE 1500 AA; 164552 MW; DCF190E7D44D89F CRC64;  
  

Query Match 9.9%; Score 86; DB 1; Length 1500;  

Best Local Similarity 23.7%; Pred. No. 22;  

Matches 27; Conservative 18; Mismatches 33; Indels 36; Gaps  
  

QY 76 SNPTQEIQQWTPEDQKAREAFRRYQEERPPETTIPPS-----SPQWKIQ 125  

DB NEPT-----PVKTDPQEPSKEETYTEKPLEPAVPAPSVENEPTPVKTPDQEPS 842  

QY 126 PGDDPLLGNSLLETHPL---XOSEPAVAVP-----IKTTPPL 158  

DB KPEPNYETKPLEPAVPAPSVENEPTPVKTPDQEPSKPEPYDPLTPPL 896  
  

RESULT 15  

EBN4_EBV STANDARD; PRT; 938 AA.  

AC P03203;  

DT 21-JUL-1986 (Rel. 01, Created)  

DT 01-AUG-1990 (Rel. 15, Last sequence update)  

DT 01-DEC-1992 (Rel. 24, Last annotation update)  

DE EBN4-4 nuclear protein (EBNA-3B).  

OS BERE7A-BERF2B.  

GN Epstein-barr virus (strain B95-8) (Human herpesvirus 4).  

OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  

OX Gammaherpesvirinae; Lymphocryptovirus.  

OC NCBI_TaxId=10377;  

RN [1]  

RP SEQUENCE FROM N.A.  

RX MEDLINE=84270657; PubMed=6087149;  

RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,  

RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Sequin C.,  

RA Tuftnell P.S., Barrell B.G.;  

RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome."  

RL Nature 310:207-211(1984).  

RN [2]  

RP CHARACTERIZATION.  

RX MEDLINE=90204695; PubMed=2157061;  

RX Kerdiles B., Wallis D., Triki H., Perricaudet M., Joab I.;  

RT "cDNA cloning and transient expression of the Epstein-Barr virus-  

RT determined nuclear antigen EBNA3B in human cells and identification  

RT of novel transcripts from its coding region.";  

J. Virol. 64:1812-1816(1990).  

RN [3]  

RP SUBCELLULAR LOCATION.  

RX MEDLINE=90266473; PubMed=21611150;  

RA Pettit L., Sample C., Kieff E.;  

RT "Subnuclear localization and phosphorylation of Epstein-Barr virus  

RT latent infection nuclear proteins.";  

VIROLOGY 176:563-574(1990).  

CC -!- FUNCTION: INVOLVED IN LATENT CYCLE.
```



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2003, 13:39:28 ; Search time 57.7782 Seconds  
(without alignments)  
371.305 Million cell updates/sec

Title: US-09-818-066-34

Perfect score: 871  
Sequence: 1 MGHPAKSMQVRIEGEGL.....PLYQSEPAVPVTKTPPLKKR 161

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08  
Maximum Match 1008

Listing first 45 summaries

Database : A\_Geneseq\_101002.\*

```
1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	871	100.0	161	18	AAW11854 Duck hepatitis B v
2	95.5	11.0	424	21	AAW11854 Arabidopsis thalia
3	95.5	11.0	559	21	AAW11854 Arabidopsis thalia
4	95.5	11.0	633	21	AAW11854 Arabidopsis thalia
5	94.5	10.8	1594	23	ABW08237 Modified human/mou
6	94	10.8	1567	23	ABW08236 Transgenic mouse N
7	93.5	10.7	906	18	AAW25681 Human c19A4C-like p
8	93.5	10.7	907	19	AAW49032 Human c19A4C-like p
9	93.5	10.7	940	19	AAW71370 Death associated p
10	93.5	10.7	940	21	AAW58970 Breast and ovarian

11	92	10.6	887	23	ABW57378
12	90	10.3	407	20	AAW93400
13	90	10.3	407	22	AAW93400
14	89.5	10.3	907	23	ABW57168
15	89	10.2	1319	22	AAW40120
16	89	10.2	1337	22	AAW07399
17	89	10.2	1342	22	ABW08234
18	88.5	10.2	586	22	AAW17059
19	88.5	10.2	1194	14	AAW42199
20	87.5	10.0	160	23	AAW21536
21	87.5	10.0	396	22	AAW38840
22	87.5	10.0	407	22	AAW23666
23	87.5	10.0	407	22	AAW40626
24	86.5	9.9	525	22	ABW44546
25	86.5	9.9	528	21	AAW33501
26	86.5	9.9	528	21	AAW33501
27	86.5	9.9	548	23	ABW97145
28	86.5	9.9	1190	22	ABW60516
29	86.5	9.9	1700	22	ABW08232
30	86	9.9	87	23	ABW33792
31	86	9.9	407	22	ABW67210
32	86	9.9	3263	22	ABW67210
33	85.5	9.8	336	19	AAW79094
34	85.5	9.8	336	23	ABW61795
35	85.5	9.8	337	20	AAW07754
36	85.5	9.8	543	22	AAW93450
37	85	9.8	139	22	AAW47508
38	85	9.8	280	14	AAW42055
39	85	9.8	377	22	AAW3643
40	85	9.8	674	22	ABW02891
41	85	9.8	743	22	ABW68417
42	84.5	9.7	301	22	ABW68068
43	84.5	9.7	520	16	AAW78185
44	84	9.6	214	22	ABW3964
45	84	9.6	417	20	AAW97698

#### ALIGNMENTS

RESULT 1	AAW11854	AAW11854 standard; Protein; 161 AA.
ID	AAW11854	
AC	AAW11854	
XX	10-MAY-1997	(first entry)
XX	Duck hepatitis B virus pre-S domain.	
DE	Duck hepatitis B virus pre-S domain.	
XX	Hepadnavirus receptor; p120; p170; vaccine; pre-S domain.	
XX	Duck hepatitis B virus.	
XX	Key	Location/Qualifiers
XX	Binding-site	87..102
XX	/label-	p170 binding site
XX	/note-	*the p170 binding site has been mapped to a major neutralising epitope of the pre-S domain (aa87-102), within which the Lys-95 and Arg-97 residues required for virion-receptor interaction*
XX	Peptide	1..102
XX	/note-	*claimed pre-S fragment (Claim 20)
XX	Peptide	1..104
XX	/note-	*claimed pre-S fragment (Claim 20)
XX	Peptide	1..126
XX	/note-	*claimed pre-S fragment (Claim 20)
XX	Peptide	1..138
XX	/note-	*claimed pre-S fragment (Claim 20)

Rat mucocardial ce  
M. tuberculosis RP  
Mycobacterium tube  
Mouse ischaemic co  
Human polypeptide  
Mouse oestrogen-re  
Novel human diago  
Novel signal trans  
HSMG1.1. Homo sa  
Human mitochondria  
Human polypeptide  
Novel human enzyme  
Human polypeptide  
Human wound healin  
Human negative elo  
CD1P-22, Incyte I  
Human tumour anti  
Drosophila melanog  
Novel human diago  
Human kinase-like  
C glutamicum prot  
Drosophila melano  
Human secreted pro  
Human polypeptide  
Human polypeptide,  
Protonibacterium  
Insecticidal/Fung  
Human extracellular  
Novel human diago  
Drosophila melano  
Drosophila melano  
Protein sequence o  
Drosophila melano  
Staphylococcus aur

FT	Peptide	25..102	"Capable of binding hepadnavirus receptor"
FT	/note=	"Claimed pre-S fragment (Claim 20)	
FT	Peptide	25..104	"Capable of binding hepadnavirus receptor"
FT	/note=	"Claimed pre-S fragment (Claim 20)	
FT	Peptide	25..126	"Capable of binding hepadnavirus receptor"
FT	/note=	"Claimed pre-S fragment (Claim 20)	
FT	Peptide	25..161	"Capable of binding hepadnavirus receptor"
FT	/note=	"Claimed pre-S fragment (Claim 20)	
FT	Peptide	42..102	"Capable of binding hepadnavirus receptor"
FT	/note=	"Claimed pre-S fragment (Claim 20)	
FT	Peptide	59..104	"Capable of binding hepadnavirus receptor"
FT	/note=	"Claimed pre-S fragment (Claim 20)	
FT	Peptide	59..126	"Capable of binding hepadnavirus receptor"
FT	/note=	"Claimed pre-S fragment (Claim 20)	
FT	Peptide	59..161	"Capable of binding hepadnavirus receptor"
FT	/note=	"Claimed pre-S fragment (Claim 20)	
FT	Peptide	71..161	"Capable of binding hepadnavirus receptor"
FT	/note=	"Claimed pre-S fragment (Claim 20)	
FT	Peptide	71..126	"Capable of binding hepadnavirus receptor"
FT	/note=	"Claimed pre-S fragment (Claim 20)	
FT	Peptide	80..104	"Capable of binding hepadnavirus receptor"
FT	/note=	"Claimed pre-S fragment (Claim 20)	
FT	Peptide	80..161	"Capable of binding hepadnavirus receptor"
FT	/note=	"Claimed pre-S fragment (Claim 20)	
FT	Peptide	87..161	"Capable of binding hepadnavirus receptor"
FT	/note=	"Claimed pre-S fragment (Claim 20)	
FT	Peptide	87..104	"Capable of binding hepadnavirus receptor"
FT	/note=	"Claimed pre-S fragment (Claim 21)	
XX			"Capable of binding hepadnavirus receptor"
PN	WO9704000-A1.		
XX			
PD	06-FEB-1997.		
XX			
PP	22-JUL-1996;	96WO-US12098.	
XX			
PR	21-JUL-1995;	95US-0001371.	
XX			
PA	(GEHO ) GEN HOSPITAL CORP.		
XX			
PI	L1 J, Tong S, Wands JR;		
XX			
DR	WPI: 1997-132572/12.		
DR	N-PSDB: MAT59583.		
XX			
PT	p170, a new avian hepadnavirus receptor - binds to pre-S domain of		
PT	duck hepatitis B virus; attenuated p170 may be used to immunise		
PT	animals against hepadnaviral infection		
XX			
PS	Claim 19; Page 87-88; 175pp; English.		
XX			
CC	The pre-S domain (AAW11854) of the duck hepatitis B virus (DHBV)		
CC	large envelope protein binds to the cellular p170 receptor (see		
CC	also AAW11851-52) at a major neutralising epitope, within which are		
CC	basic amino acids (Lys-95, Arg-97) required for virion-receptor		
CC	interaction. Peptide sequences encompassing Lys-95 and Arg-97		
CC	can act as p170 ligands. Such peptides can be used to reduce		
CC	the level of hepadnaviral infection in an animal. The pre-S		
CC	domain (esp. attenuated), and nucleic acids encoding it, can also		

CC	be used as vaccines to immunise animals against hepadnavirus infection.
XX	
XX	
SQ	Sequence      161 AA:
	Query Match                          100.0%; Score 871; DB 18; Length 161;
	Best Local Similarity    100.0%; Pred. No. 9,4e-76;
	Matches    161; Conservative    0; Mismatches    0; Indels        0; Gaps        0;
OY	1 MGCHFAKMDVRRIGSGELLNQLAGRMIPKGTLMWSGKFPLDVLHVQTMEEINTLQ 60
DB	1 MGCHFAKMDVRRIGSGELLNQLAGRMIPKGTLMWSGKFPLDVLHVQTMEEINTLQ 60
OY	61 NCGAPACAGRVRVGSNTPTPEIPOPWTPEEDOKAREAFRYYOEERPETTTIPSSPP 120
DB	61 NCGANPAGAGRVGVLSNTPTEIPOPWTPEEDOKAREAFRYYOEERPETTTIPSSPP 120
OY	121 OWKLOPGDDPLIGNOSLLETHPHLYOSEPAVPVIKTPLKKK 161
DB	121 OWKLOPGDDPLIGNOSLLETHPHLYOSEPAVPVIKTPLKKK 161
RESULT 2	
AAG31344	
ID	AAG31344 standard; Protein; 424 AA.
XX	
AC	AAG31344;
XX	
DT	17-OCT-2000 (first entry)
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 37626.
XX	
KW	Protein identification; signal transduction pathway; metabolic pathway;
RW	hybridisation assay; genetic mapping; gene expression control; promoter;
XX	termination sequence.
XX	
CS	Arabidopsis thaliana.
PN	EP1033405-A2.
XX	
PD	06-SEP-2000.
XX	
PF	25-FEB-2000; 2000EP-0301439.
XX	
PR	25-FEB-1999; 99US-0121825.
PR	05-MAR-1999; 99US-0123180.
PR	09-MAR-1999; 99US-0123548.
PR	23-MAR-1999; 99US-0125788.
PR	25-MAR-1999; 99US-0126264.
PR	29-MAR-1999; 99US-0126785.
PR	01-APR-1999; 99US-0127462.
PR	06-APR-1999; 99US-0128234.
PR	08-APR-1999; 99US-0128714.
PR	16-APR-1999; 99US-0129845.
PR	19-APR-1999; 99US-0130077.
PR	21-APR-1999; 99US-0130449.
PR	23-APR-1999; 99US-0130510.
PR	23-APR-1999; 99US-0130891.
PR	28-APR-1999; 99US-0131449.
PR	30-APR-1999; 99US-0132048.
PR	30-APR-1999; 99US-0132407.
PR	04-MAY-1999; 99US-0132484.
PR	05-MAY-1999; 99US-0132485.
PR	06-MAY-1999; 99US-0132486.
PR	06-MAY-1999; 99US-0132487.
PR	07-MAY-1999; 99US-0132863.
PR	11-MAY-1999; 99US-0134256.
PR	14-MAY-1999; 99US-0134218.
PR	14-MAY-1999; 99US-0134219.
PR	14-MAY-1999; 99US-0134221.
PR	14-MAY-1999; 99US-0134370.
PR	18-MAY-1999; 99US-0134768.
PR	19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139753.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139859.  
PR 23-JUN-1999; 99US-0140353.  
PR 24-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145087.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145224.  
PR 23-JUL-1999; 99US-0145276.  
PR 26-JUL-1999; 99US-01452913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149375.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 13-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.

PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 11.0%; Score 95.5; DB 21; Length 424;  
Best Local Similarity 24.5%; Pred. No. 0.86;  
Matches 40; Conservative 15; Mismatches 45; Indels 63; Gaps 10;

QY 21 LNLQARMLPKGTLT--WSGKPFLLDVLVDH-----VQWMEINTLQNGAMPA 67  
Db 264 VQELLK---TGNVTSFW-----MDHITHCVSTPSVEENATREAVYNLQ---WPP 310  
QY 68 GAGRAY-----GLSNPTQEIROPQWTPREDQKAREAFRRYQERPPETTTIP 115  
Db 311 NGRHLIAEFVRAEVEKLEAPLP--PQPHQPOAQTLSR-----PPPTALP 356  
QY 116 PSSPQWKLQPGDDPLGNOSLETHPLQOSEPAVVIKTPPL 158  
Db 357 P--P-----PLAKPPHYVERLPLPPPIAEPEDEPPI 388

RESULT 3  
AAG31343  
ID AAG31343 standard; Protein; 559 AA.

AC AAG31343;  
XX  
DT 17-OCT-2000 (first entry)  
XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 37625.

XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

XX Arabidopsis thaliana.

OS EP1033405-A2.

PN  
XX  
PD 06-SEP-2000.

XX  
PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135123.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.

PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0149930.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153738.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.

PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161921.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
PR 29-OCT-1999; 99US-0162142.

Query Match 11.0%; Score 95.5; DB 21; Length 559;  
Best Local Similarity 24.5%; Pred. No. 1.2;  
Matches 40; Conservative 15; Mismatches 45; Indels 63; Gaps 10;

QY 21 LNQLAGRMIPKGLT-NSGKFTLDHVDH-----VQTMETINTLQNGANPA 67  
Db 399 VOELLGK---TGNVTSFM-----MDHIKTHCYVSYSVEEAAATREAVYNIQ---MP 445  
QY 68 GAGRRV-----GLSNPTQELIPQWPTEEDQKAEARRRYQGEPPETITP 115  
Db 446 NGRHLIAEFVRAEEVKELEAPL---PQPHOPOAOITLSR-----PPPTALP 491  
QY 116 PSSPPQWKLPQDDPLLNQSLLETHPLPYQSEPAVPVKTPTPL 158  
Db 492 P--PP-----PLAKPPHYVERLPLPPPPPIAPEQDEPPI 523

RESULT 4  
AAG31342  
ID AAG31342 standard; Protein: 633 AA.  
XX  
AC AAG31342;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SQ ID NO: 37624.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KM hydridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PE 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.

PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 25-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151308.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 22-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 22-SEP-1999; 99US-0155186.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.



PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161921.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match  
Best Local Similarity 24.5%; Pred. No. 1.4;  
Matches 40; Conservative 15; Mismatches 45; Indels 63; Gaps 10;

OY 21 LNLQLAGRIIPKGLT--NSGKPTLDHVDH-----VQMEIINTLQNGAMPA 67  
DB 473 VOELLGK--TGNVTSFW-----MDHITKHCVSVSYVEEAATREAVYNIQ---WPP 519  
OY 68 GAGRRV-----GLSNPTQETIPQWPTPEEDOKAREARRRQGEERPPETTTIP 115  
DB 520 NCGRHLLAEFVRAEEVKEKLEAPLP---PQPHQPOAQTLSR-----PPPTALP 565  
OY 116 PSSPPQWKLPQGDPLGNSLLETHPIYQSEPAVPVITKPPPL 158  
DB 566 P--PP-----PLAKPPHYVERLPLPPPIAPAEQGEPI 597

## RESULT 5

ABB08237  
ID ABB08237 standard; Protein: 1594 AA.

XX ABB08237;

XX 15-APR-2002 (first entry)

XX Modified human/mouse Bcr-Abl fusion protein #2.

XX Human; mouse; Bcr-Abl; cancer; cytostatic; immunosuppressive; leukaemia;  
XX orthologous bone marrow transplantation.

XX Chimeric - Homo sapiens.

XX Chimeric - Mus musculus.

XX WO200200024-A1.

XX 03-JAN-2002.

XX 29-JUN-2001; 2001WO-US20602.

XX 30-JUN-2000; 2000US-215595P.

XX (REGC ) UNIV CALIFORNIA.

XX Wang JYJ, Vigneri P;

XX WPI: 2002-139846/18.

XX N-PSDB; ABA96147.

XX Killing cancer cells for treating leukaemia, involves causing  
XX accumulation of activated Bcr-Abl in the nucleus to induce apoptosis -

XX Example 10; Page 54-61; 63pp; English.

XX The sequence represents a human/mouse Bcr-Abl fusion protein. The  
XX invention relates to a novel method for killing cancer cells, comprising  
XX causing an accumulation of activated Bcr-Abl in the nucleus, to induce  
XX apoptosis. The method has cytostatic and immunosuppressive activity. The  
XX method is useful for killing cancer cells, and in the treatment of  
XX cancers including leukaemia. The invention also supplies a method useful  
XX for purging bone marrow to allow for orthologous bone marrow  
XX transplantation, and as a therapeutic option for chronic myelogenous  
XX leukaemia.

XX Sequence 1594 AA;

Query Match  
Best Local Similarity 10.8%; Score 94.5; DB 23; Length 1594;  
Matches 35; Conservative 20; Mismatches 42; Indels 39; Gaps 7;

OY 53 MEIINTLQNGAMPAGACRVGLSNPTQETIPQWPT-----PEEDOKAREARRR- 102  
DB 1356 MDAVNTDPTK-AGPGE-----GLRRPVPPSVKPPQSTGGGDKPKRRKRVPPKRRKRV 1410  
OY 103 -----YQBE-----RPPETTTIPSSPPQWKLPQGDPLIGNO--SLETH 141  
DB 1411 GSGTGGGDYKDDDKGGGGSTAKPPTSVTSVP---STRAPASPLAGDQGPSSAFT 1467  
OY 142 PLYQSEPAVPVITKPP 157  
DB 1468 PLISTRVSLRKTRQPP 1483

## RESULT 6

ABB08236  
ID ABB08236 standard; Protein: 1567 AA.

XX ABB08236;

XX 15-APR-2002 (first entry)

XX Modified human/mouse Bcr-Abl fusion protein #1.

XX Human; mouse; Bcr-Abl; cancer; cytostatic; immunosuppressive; leukaemia;  
XX orthologous bone marrow transplantation.

XX Chimeric - Homo sapiens.

XX Chimeric - Mus musculus.

XX WO200200024-A1.

XX 03-JAN-2002.

XX 29-JUN-2001; 2001WO-US20602.

XX 30-JUN-2000; 2000US-215595P.

XX (REGC ) UNIV CALIFORNIA.

XX Wang JYJ, Vigneri P;

XX WPI: 2002-139846/18.

XX N-PSDB; ABA96146.

XX Killing cancer cells for treating leukaemia, involves causing  
XX accumulation of activated Bcr-Abl in the nucleus to induce apoptosis -

XX Disclosure; Page 40-47; 63pp; English.

XX The sequence represents a human/mouse Bcr-Abl fusion protein. The  
XX invention relates to a novel method for killing cancer cells, comprising  
XX causing an accumulation of activated Bcr-Abl in the nucleus, to induce  
XX apoptosis. The method has cytostatic and immunosuppressive activity. The  
XX method is useful for killing cancer cells, and in the treatment of  
XX cancers including leukaemia. The invention also supplies a method useful  
XX for purging bone marrow to allow for orthologous bone marrow  
XX transplantation, and as a therapeutic option for chronic myelogenous  
XX leukaemia.

XX Sequence 1567 AA;

Query Match  
Best Local Similarity 10.8%; Score 94; DB 23; Length 1567;  
Matches 33; Conservative 14; Mismatches 46; Indels 20; Gaps 6;

OY 53 MEIINTLQNGAMPAGACRVGLSNPTQETIPQWPT-----PEEDOKAREARRQEE 106  
DB 1356 MDAVNTDPTK-AGPGE-----GLRRPVPPSVKPPQSTGGGDYKDDDKGGGG-----STA 1406

QY 107 RPEPTTIPSSPPQMKLOPQDDPLGNO--SLETHPLXQSEPAVVIKTPP 157  
 Db 1407 KPGPTIPSPVSTP--STAPAPSLAGDQDPSSAAFIPLISTVSLKTRQPP 1456

RESULT 7  
 AAM25681  
 ID AAM25681 standard; Protein; 906 AA.  
 AC AAM25681;  
 XX  
 DT 31-MAR-1998 (first entry)  
 DE Transgenic mouse NTAl polypeptide.  
 XX  
 KW RNA editing; mooring primer; screening; detection; editing enzyme;  
 KM obesity; APOBEC-1; therapeutic; NTAl; transgenic.  
 XX  
 OS Synthetic.  
 OS Homo sapiens.  
 OS Mus sp.

Key Location/Qualifiers  
 FT 1..906  
 FT Protein /label= "NTAl  
 FT /note= "partial coding sequence"

MO9732032-A1.  
 PD 04-SEP-1997.  
 XX  
 PF 28-FEB-1997; 97WO-US03189.  
 XX  
 PR 01-MAR-1996; 96US-0609230.  
 XX  
 PA (REGC ) UNIV CALIFORNIA.  
 XX  
 PI Innerarity TL, Qian X, Yamanaka S;  
 XX  
 DR WPI: 1997-448694/41.  
 DR N-PSDB; AAT86087.  
 XX  
 PT Detection of mRNA targets for editing enzymes - by amplifying RNA  
 PT from tissues by RT-PCR using mooring primers to identify genes  
 PT responsible for non-wild type phenotype(s)  
 XX  
 PS Example 4; Figure 5; 36pp; English.

This sequence represents a novel target of the apob mRNA-editing  
 CC enzyme catalytic polypeptide #1, APOBEC-1, (i.e. NTAl) which is used  
 CC as an example of a novel method of detecting mRNA candidates for editing  
 CC in a tissue. The method involves the generation of cDNA from RNA in the  
 CC tissue using one or more mooring primers and reverse transcriptase (RT)  
 CC and is amplified by PCR using one or more mooring primers and a 5'  
 CC arbitrary primer. The method has been used for identification of an  
 CC obesity gene e.g. APOBEC-1 or an oncogene in an animal and also for the  
 CC identification of non-wild type phenotypes. The identification of such  
 CC mRNA's provides insight into the role of the genes encoding them, in  
 CC growth, differentiation, or lipid accumulation, and also into the  
 CC possible role of an RNA editing enzyme or polypeptide in causing obesity  
 CC or as an oncogene. The identified polypeptides can be used as therapeutic  
 CC reagents in situations where the edited version of the polypeptide  
 CC results in a pathological state. Additionally, such polypeptides or the  
 CC edited versions can be used in the preparation of antibodies for  
 CC therapeutic use.

Sequence 906 AA:  
 XX  
 SQ

Query Match 10.7%; Score 93.5; DB 18; Length 906;  
 Best Local Similarity 25.3%; Pred. No. 3.5;  
 Matches 44; Conservative 30; Mismatches 59; Indels 41; Gaps 10;

QY 1 MGCHPAKSMQVRRIRGESEIL-----LNOLAGNMIRPGITLWSGKFPPTLDHVDHYQTM 54

Db 397 MGRH--RSNOLENGCHGHPPTQSOFGEMGFMKSQGIS-----OLYHNSOG 444  
 QY 55 EINTLONOGAMPAGARRVGLSNPTPOEIRPOPWTPPE--DQKAREAFRRYOERP--PE 110  
 Db 445 LLSOLOGQS-----KDMPPRFSSKKGOLNADSLRRAQSLMKKNQVPLKLPQ 492  
 QY 111 TTTTIPSS-PPQMKLOP-GDDPLIGNQSLLETH-PLYOSEPAVVIKTPPLKKR 161  
 Db 493 ITMIPSAQPPRTQTPPLGOTPOLG----LKTNPPLIQEKPAKTSKKPPSKKE 542

RESULT 8  
 AAM49032  
 ID AAM49032 standard; Protein; 907 AA.  
 AC AAM49032;  
 XX  
 DT 07-OCT-1998 (first entry)  
 DE Human eIF4G-like protein (p97).  
 XX  
 KW Human p97-like subunit protein; hprt1; transcription; apoptosis;  
 KM p97; antagonist; agonist; human eIF4G-like protein.  
 XX  
 OS Homo sapiens.

Key Location/Qualifiers  
 FH 788..802  
 FH Peptide /note= "INT peptide used to raise anti-INT  
 FT /note= "INT peptide used to raise anti-INT  
 FT antibodies"

WO9825957-A2.  
 PN 18-JUN-1998.  
 XX  
 PD 12-DEC-1997; 97WO-US22664.  
 XX  
 PR 13-DEC-1996; 96US-0033151.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PA (UYMC-) UNIV MCGILL.  
 XX  
 PI Imataka H, Methot N, Olsen HS, Rom E, Ruben SM,  
 PI Sorenberg N;  
 XX  
 DR WPI: 1998-348451/30.  
 DR N-PSDB; AAV32791.  
 XX  
 PT Polynucleotides encoding human hprt1 and p97 - useful for treating  
 PT disease states associated with apoptosis  
 XX  
 PS Claim 1; Fig 2A-2E; 110pp; English.

The present sequence represents a DNA sequence encoding a human  
 CC eIF4G-like protein, referred to as p97. The p97 DNA sequence was  
 CC isolated from human embryo brain cDNA library. p97 is involved in  
 CC initiation of transcription. The invention claims for the hprt1  
 CC (AAM49031) and p97 proteins. The hprt1 and p97 proteins are claimed  
 CC to be useful in screening for (ant)agonists of hprt1 and/or p97  
 CC activity. These proteins are also claimed to be useful for treating a  
 CC disease state associated with apoptosis. Anti-hprt1 and anti-p97  
 CC antibodies are useful for identifying and isolating the hprt1 and p97  
 CC proteins respectively.

Sequence 907 AA:  
 XX  
 SQ

Query Match 10.7%; Score 93.5; DB 19; Length 907;  
 Best Local Similarity 25.3%; Pred. No. 3.5;  
 Matches 44; Conservative 30; Mismatches 59; Indels 41; Gaps 10;

QY 1 MGCHPAKSMQVRRIRGESEIL-----LNOLAGNMIRPGITLWSGKFPPTLDHVDHYQTM 54

Db 398 MGRH--RSNOLFNGHGHIMPTOSQFEGMGKFKSGSL-----QLYHNSOG 445

QY 55 EINTLQNGAMPAGARRVGLSNPTPOEIPPOWTPPE-DOKARAFRRYOERP---PE 110

Db 446 LLSQLQSGS-----KDMPPRSKKGQLNADISLRPAQSFIMNNQVPLKLPQ 493

QY 111 TTTTPPS-PPQWKLP-GDDPLLGNSLLETH-PLYSEPAVPIKTPPLKK 161

Db 494 ITMTPSAOPPRQTPTPLGQTPQLG---LKTNPPLIOEKPAKTSKKPPSKKE 543

RESULT 9

AAW71370

ID AAW71370 standard; Protein: 940 AA.

AC AAW71370;

XX 12-JAN-1999 (first entry)

DE Death associated protein (DAP)-5.

XX Death associated protein: DAP-5; cell death; tumour cell;

KW metastatic activity; cancer; psoriasis; autoimmune disease;

KW programmed cell death; degenerative neurological disease;

KW Alzheimer's.

XX Homo sapiens.

OS

XX MO8939429-A2.

PN

XX 11-SEP-1998.

PD

XX 03-MAR-1998; 98MO-IL00102.

PF

XX 03-MAR-1997; 97US-0810712.

PR

XX (YEDA ) YEDA RES & DEV CO LTD.

PA

XX Kimchl A;

PI

XX WPI: 1998-520781/44.

DR

XX N-PSDB; AAV60293.

DR

XX

XX New isolated death associated protein nucleic acids - used for the

PT diagnosis and treatment of disorders associated with programmed cell

PT death, e.g. cancers, autoimmune disease or neurological disease

PS Disclosure; Fig 15; 157pp; English.

XX

XX The present sequence represents a death associated protein (DAP)-5.

CC The DAP genes and proteins are used for promoting death of normal or

CC tumour cells, and for suppressing the metastatic activity of tumour

CC cells. They can be used in the treatment of diseases or disorders

CC associated with uncontrolled pathological growth, e.g. cancer,

CC psoriasis, autoimmune diseases and others. Agents which antagonise,

CC inhibit or neutralize DAP products are used for protecting cells from

CC programmed cell death. In this case they can be used for the treatment

CC of degenerative neurological diseases, e.g. Alzheimer's, prevention of

CC death of T cells in AIDS patients, prevention of rejection associated

CC cell death in transplants, and protection of normal cells from

CC the cytotoxic effects of anti-cancer therapies.

XX

SQ Sequence 940 AA;

Query Match 10.7%; Score 93.5; DB 19; Length 940;

Best Local Similarity 25.3%; Pred. No. 3.6;

Matches 44; Conservative 30; Mismatches 59; Indels 41; Gaps 10;

QY 1 MGOIPAKSMQVRIEGEIT-----LNLQGRMIPKGLTWGKFTPLDHLVDVOTME 54

Db 431 MGRH--RSNOLFNGHGHIMPTOSQFEGMGKFKSGSL-----QLYHNSOG 478

QY 55 EINTLQNGAMPAGARRVGLSNPTPOEIPPOWTPPE-DOKARAFRRYOERP---PE 110

Db 479 LLSQLQSGS-----KDMPPRSKKGQLNADISLRPAQSFIMNNQVPLKLPQ 526

QY 111 TTTTPPS-PPQWKLP-GDDPLLGNSLLETH-PLYSEPAVPIKTPPLKK 161

Db 527 ITMTPSAOPPRQTPTPLGQTPQLG---LKTNPPLIOEKPAKTSKKPPSKKE 576

RESULT 10

AAB58970

ID AAB58970 standard; Protein: 940 AA.

AC AAB58970;

XX 27-MAR-2001 (first entry)

DE Breast and ovarian cancer associated antigen protein sequence SEQ ID 678.

XX Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;

KW neurotropic; neuroprotective; antiviral; antiallergic; hepatotropic;

KW antidiabetic; antiinflammatory; antitumor; vulnery; anticonvulsant;

KW antibacterial; antifungal; antiparasitic; cardiant; immune disorder;

KW Addison's disease; allergy; autoimmune haemolytic anaemia;

KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;

KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;

KW cardiovascular disorder; wound healing; neurological disease.

XX Homo sapiens.

OS

XX MO200055173-A1.

PN

XX 21-SEP-2000.

PD

XX 08-MAR-2000; 2000WO-US05881.

PF

XX 12-MAR-1999; 99US-0124270.

PR

XX (HUMA-) HUMAN GENOME SCI INC.

PA

XX Rosen CA, Ruben SM.

PI

XX WPI: 2000-611515/58.

DR

XX N-PSDB; AAF21873.

DR

XX

XX New human breast and ovarian cancer associated gene sequences and the

PT polypeptides encoded by these genes, useful in the prevention,

PT treatment and diagnosis of cancer, immune disorders, cardiovascular

PT disorders and neurological diseases -

PS Claim 11; Page 1128-1132; 1299pp; English.

XX

XX Sequences AAF21614 - AAF22031 represent DNA sequences encoding human

CC proteins AAB58711 - AAB59128. The DNA and protein sequences are

CC associated with breast and ovarian cancer. Included in the invention are

CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the

CC isolation and characterisation of the DNA and protein sequences of the

CC invention. The breast and ovarian cancer associated DNA, protein, agonist

CC or antagonist sequences exhibit cytostatic, immunosuppressive;

CC neurotropic; neuroprotective; antiviral; antiallergic; hepatotropic;

CC antidiabetic; antiinflammatory; antitumor; vulnery; anticonvulsant;

CC antibacterial; antifungal; antiparasitic and cardiant activity. The

CC polynucleotide and protein sequences are used in the diagnosis of cancer,

CC particularly breast and ovarian cancer. The nucleic acid sequences,

CC proteins, agonists and antagonists may also be used in the diagnosis,

CC prevention and treatment of immune disorders e.g. Addison's disease,

CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,

CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid

CC arthritis and ulcerative colitis; cardiovascular disorders such as

CC myocardial ischemias; wound healing; neurological diseases such as

CC cerebral anoxia and epilepsy; and infectious diseases.

XX

SQ Sequence 940 AA;



OY 11 VRIIE-GGEILLN---OLAGRMIPKGLTWS---GKFPILDHVLHVQTM---EEINT 58  
ID AAB52470 standard; protein: 407 AA.  
DB 47 VARESGGNMISINTNGNYLGIGLQFTQS--TWAHAGGGEFAPSQAQLASRQQLAVGERVLA 104  
OY 59 LQNGCAMPAGRGVGLSNPTPOEIPQWTPPEEQKAREFRKROERPPETTTIPSS 118  
DB 105 TQGRGAMPV-CGR--GLSNATPREVLPAASAMADAPLDAAV-----NGEPAPLAPPAD 155  
OY 119 P-POWKLOPQDDPLGNOSLLETHPLVQSEPAVPIKTRP 157  
DB 156 PAPPELALANDLPA-----PLGEPLPAAPADPAP 185

RESULT 13  
AAB52470  
ID AAB52470 standard; protein: 407 AA.  
AC AAB52470;  
DT 23-FEB-2001 (first entry)  
DE Mycobacterium tuberculosis secreted protein #35.  
KW Mycobacterium tuberculosis secreted protein; MTSP; vaccine.  
OS Mycobacterium tuberculosis.  
PN WO200066143-A1.  
PD 09-NOV-2000.  
PE 04-MAY-2000; 2000WO-US12197.  
PR 04-MAY-1999; 99US-0132479.  
PR 04-MAY-1999; 99US-0132503.  
PA (PUBL-) PUBLIC HEALTH RES INST NEW YORK.  
PI Gennaro ML, Gomez MJ;  
PT WPI: 2001-007151/01.  
PT Novel Mycobacterium tuberculosis secreted polypeptides and  
PT polynucleotides useful in diagnosis, treatment and prophylaxis of  
PT tuberculosis -  
PS Claim 11; Fig 1; 60pp; English.  
XX The present invention relates to Mycobacterium tuberculosis secreted  
XX proteins (MTSP), where the polypeptide has M. tuberculosis specific  
XX antigenic and immunogenic properties. Compositions of the invention may  
XX be useful for diagnosing Mycobacterium tuberculosis infection and as a  
XX vaccine against M. tuberculosis infection.  
SQ Sequence 407 AA;  
Query Match 10.3%; Score 90; DB 22; Length 407;  
Best Local Similarity 27.5%; Pred. No. 2.7;  
Matches 44; Conservative 14; Mismatches 68; Indels 34; Gaps 10;

RESULT 14  
AAB57168  
ID AAB57168 standard; protein: 907 AA.  
AC AAB57168;  
DT 07-MAR-2002 (first entry)  
DE Mouse ischaemic condition related protein sequence SEQ ID NO:408.  
KW Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;  
KW vasospastic ischaemia; ischaemic condition; ischaemic disease.  
OS Mus musculus.  
PN WO200188188-A2.  
PD 22-NOV-2001.  
PE 18-MAY-2001; 2001WO-JP04192.  
PR 18-MAY-2000; 2000JP-0145977.  
PR (UYNI-) UNIV NIHON SCHOOL JURIDICAL PERSON.  
PA Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;  
PI WPI: 2002-034733/04.  
DR N-PSDB: ABI99460.  
DR Examining the ischemic condition (e.g. occlusive ischemia) by measuring  
DR expression levels of particular genes defined in the specification or  
DR by determining the expression profile of a gene group comprising these  
DR genes -  
PS Claim 2; Page 1113-1117; 2690pp; English.  
XX The present invention describes a method for examining ischemic  
XX conditions, comprising measuring the expression levels of particular  
XX genes (I) in a test sample or determining the expression profile of a  
XX gene group in the sample comprising genes selected from (I). The method  
XX is useful for examining the ischemic condition (e.g. compressive  
XX ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring  
XX expression levels of particular genes (ABI99202 to ABI99912, encoding  
XX the protein sequences in ABB57020 to ABB57374) or by determining the  
XX expression profile of a gene group comprising these genes. The  
XX expression levels or expression profiles produced by these genes are  
XX used as an indicator when screening for ischaemic condition-improving  
XX drugs or therapeutics for ischaemic diseases. ABI9913 and ABI9914  
XX represent PCR primers for a mouse ischaemic condition related sequence,  
XX which are used in the exemplification of the present invention.  
SQ Sequence 907 AA;  
Query Match 10.3%; Score 89.5; DB 23; Length 907;  
Best Local Similarity 24.4%; Pred. No. 8.4;  
Matches 44; Conservative 31; Mismatches 52; Indels 53; Gaps 11;

RESULT 15

ID	AA#0120	standard; Protein; 1319 AA.
XX	AA#0120	
AC	AA#0120;	
DT	22-OCT-2001	(first entry)
DE	Human polypeptide SEQ ID NO 3265.	
XX		
KW	Human; nocitropic; immunosuppressant; cytostatic; gene therapy; cancer;	
KW	peripheral nervous system; neuropathy; central nervous system; CNS;	
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;	
KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;	
KW	chemokine; thrombolytic; drug screening; arthritis; inflammation;	
KW	leukaemia.	
XX		
OS	Homo sapiens.	
XX		
PN	MO20015312-A1.	
XX	26-JUL-2001.	
PD		
XX	26-DEC-2000; 2000WO-US34263.	
PF		
XX	21-JAN-2000; 2000US-0488725.	
PR	25-APR-2000; 2000US-0552317.	
PR	09-JUL-2000; 2000US-0596042.	
PR	19-JUL-2000; 2000US-0620312.	
PR	03-AUG-2000; 2000US-0653450.	
PR	14-SEP-2000; 2000US-0662191.	
PR	19-OCT-2000; 2000US-0693036.	
PR	29-NOV-2000; 2000US-0727344.	
XX		
PA	(HXSE-) HXSEQ INC.	
XX		
PI	Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;	
PI	Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;	
PI	Zhao QA, Zhou P, Goodrich R, Drmanac RT;	
XX		
DR	WPI: 2001-442253/47.	
DR	N-PSDB: AAI59276.	
XX		
PT	Novel nucleic acids and polypeptides, useful for treating disorders	
PT	such as central nervous system injuries -	
XX		
PS	Example 5; SEQ ID NO 3265; 10078bp; English.	
XX		
CC	The invention relates to human nucleic acids (AA15798-AA161369) and	
CC	the encoded polypeptides (AA#38642-AA#42213) with nocitropic,	
CC	immunosuppressant and cytostatic activity. The polynucleotides are useful	
CC	in gene therapy. A composition containing a polypeptide or polynucleotide	
CC	of the invention may be used to treat diseases of the peripheral nervous	
CC	system, such as peripheral nervous injuries, peripheral neuropathy and	
CC	localised neuropathies and central nervous system diseases, such as	
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic	
CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the	
CC	utilisation of the activities such as: Immune system suppression,	
CC	Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic	
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,	
CC	assays for receptor activity, arthritis and inflammation, leukaemias and	
CC	C.N.S disorders.	
CC	Note: The sequence data for this patent did not form part of the printed	
CC	specification.	
XX		
SEQ	Sequence 1319 AA;	
Query Match	10.2%;	Score 89; DB 22; Length 1319;
Best Local Similarity	28.6%;	Pred. No. 15;
Matches 32; Conservative	9;	Mismatches 41; Indels 30; Gaps 4.
64	AMPACAGRVGSGN---PTPOEIPDPQWPEEDOKARAEFRVQ-----	104
276	AMPRAAG-MVGLSKMHAQDPQDQPPQDQDQGVFFEFKSGARKMPVGLSPVSGR	334

```

QY      105 -----EERPPTTTTIPSSSPQWKLQPGDDPLGNOSLLETHPLYQSEPA 149
      :: || || || | : | : || || || ||
Db      335 HPLMQPPQQAAPPQQQPPQPPQQQPPPPGGLVLRQNSLPACAP---SAPA 383

```

Search completed: January 2, 2003, 13:44:23  
Job time : 59.7782 secs

```
Ox      105 -----ERPPTTTTIPSSPOMKIQGDDPLIGNOSLLETHPLYOSEPA 149
          :: ||         || ||| : |         | : ||         | |||
Db      335 HPLMQPPQQAPPPPQQQFPQQFPQQQPPPPPGLLVRONSLPACAF---SAPA 383
```

Search completed: January 2, 2003, 13:44:23  
Job time : 59.7782 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2003, 13:42:28 ; Search time 46.0927 Seconds  
(without alignments)  
719.714 Million cell updates/sec

Title: US-09-818-066-34

Perfect score: 871

Sequence: 1 MGQHPAKSMVRRIEGGEIL.....PLYQSEPAVPVKTPLKKK 161

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL21:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_rvirus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	871	100.0	329	12	O92935 duck hepati
2	866	99.4	364	12	O66399 duck hepati
3	857	98.4	328	12	O8QX1 duck hepati
4	785.5	90.2	330	12	O72885 duck hepati
5	781.5	89.7	329	12	O91HP5 duck hepati
6	779.5	89.5	330	12	O66405 duck hepati
7	779.5	89.5	366	12	O66404 duck hepati
8	724	83.1	329	12	O9WFA3 snow goose
9	724	83.1	329	12	O9WFA9 snow goose
10	724	83.1	329	12	O9WFB3 snow goose
11	717	82.3	329	12	O9WFB6 snow goose
12	714	82.0	329	12	O9WFA6 snow goose
13	602.5	69.2	327	12	O67852 duck hepati
14	444	51.0	337	12	O8UYX6 stork hepat
15	444	51.0	337	12	O8UYX4 stork hepat
16	439	50.4	337	12	O8UYX0 stork hepat

17	439	50.4	337	12	O8UYX8
18	101	11.6	1678	11	O924C5
19	98.5	11.3	418	16	O9A4Y7
20	98	11.3	1259	4	O8WXX7
21	96	11.0	437	4	O96S57
22	96	11.0	466	4	O96ANA
23	96	11.0	482	16	O8YUM8
24	96	11.0	502	4	O8TBT1
25	96	11.0	548	4	O9BTI7
26	96	11.0	559	4	O9BYJ9
27	95.5	11.0	315	11	O55150
28	95.5	11.0	633	10	O65655
29	95.5	11.0	1339	11	O35788
30	94.5	10.8	326	11	O91WA8
31	94.5	10.8	542	12	O84357
32	94	10.8	1260	4	O9UGY9
33	93.5	10.7	907	4	P78344
34	93.5	10.7	907	6	P79398
35	92.5	10.6	269	10	O9FUR7
36	92.5	10.6	989	11	O9JLE9
37	92	10.6	539	10	O9M287
38	92	10.6	981	15	O92809
39	91.5	10.5	299	10	O49201
40	91.5	10.5	299	10	O94G51
41	91.5	10.5	1386	4	O9C0A3
42	91	10.4	197	4	O9NX79
43	90.5	10.4	926	3	O13305
44	90	10.3	407	16	O53879
45	89.5	10.3	907	11	O62448

#### ALIGNMENTS

##### RESULT 1

O92935 ID O92935 PRELIMINARY; PRT; 329 AA.  
AC O92935;  
DT 01-NOV-1998 (TREMBLrel. 08, Created)  
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE PreS antigen.  
OS Duck hepatitis B virus (DHBV).  
OC Viruses; Retroid viruses; Hepadnaviridae; Avihepadnavirus.  
OX NCBI\_TaxID=12639;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ALBERTA;  
RA Fischer K.P., Stickney J., Tipples G.A., Tyrrell D.L.J.;  
RT "Cloning, sequencing and sequence comparison of a Canadian isolate of  
RT duck hepatitis B virus."  
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF047045; AAC06355.1;  
DR InterPro: IPR000349; Hepvir\_surflag.  
DR Pfam: PF00695; WMSA; 1.  
SQ SEQUENCE 329 AA; 36361 MW; 46E4ACAF4995147 CRC64;

Query Match 100.0%; Score 871; DB 12; Length 329;  
Best Local Similarity 100.0%; Pred. No. 1.8e-68;  
Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MGQHPAKSMVRRIEGGEILLNQLAGRMIPKGTLTWSGKFTLDHVLHVQTMEEINTLQ 60  
Db 2 MGQHPAKSMVRRIEGGEILLNQLAGRMIPKGTLTWSGKFTLDHVLHVQTMEEINTLQ 61  
Qy 61 NOGAWPAGAGRVGLSNTPQEIPOQWTPPEQDKARAFRYQERPEPTTTPPSPP 120  
Db 62 NOGAWPAGAGRVGLSNTPQEIPOQWTPPEQDKARAFRYQERPEPTTTPPSPP 121  
Qy 121 QWKLPQGGDDPLLGNQSLLETHPLVQSEPAVPVKTPLKKK 161  
Db 122 QWKLPQGGDDPLLGNQSLLETHPLVQSEPAVPVKTPLKKK 162

```
RESULT 2
Q66399
ID Q66399 PRELIMINARY; PRT; 364 AA.
AC Q66399;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Surface antigen.
OS Duck hepatitis B virus (DHBV).
OC Viruses; Retrovird viruses; Hepadnaviridae; Avihepadnavirus.
OX NCBI_TaxID=12639;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=INDIAN;
RA Munshi A., Panda S.K.;
RT "Cloning sequencing and sequence comparison of the indian isolate.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; X74623; CAA52699.1; -.
DR InterPro; IPR000349; Hepvir_surfa.
DR Pfam; PF00695; vmsa; 1.
SQ SEQUENCE 364 AA; 40385 MW; E2E27FB4E4775C19 CRC64;

Query Match 99.4%; Score 856; DB 12; Length 364;
Best Local Similarity 99.4%; Pred. No. 5.4e-68;
Matches 160; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MQGHPAKSMVRRIGEGELLNQLAGRMIPKGTLTWSGKFFPTLDHVDHVQMEINTLQ 60
DB 37 MQGHPAKSMVRRIGEGELLNQLAGRMIPKGTLTWSGKFFPTLDHVDHVQMEINTLQ 96
QY 61 NOGAMPAGARRVGLSNPTTPOIQTPEEDQKAREAFRRYQERPPETTTIPSSPP 120
DB 97 NOGAMPAGARRVGLSNPTTPOIQTPEEDQKAREAFRRYQERPPETTTIPSSPP 156
QY 121 QWKLPQDDPLNGSLLETHPLYSQSEPAVPVTKPPLKKK 161
DB 157 QWKLPQDDPLNGSLLETHPLYSQSEPAVPVTKPPLKKK 197

RESULT 3
Q80QX1
ID Q80QX1 PRELIMINARY; PRT; 328 AA.
AC Q80QX1;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Presurface protein.
OS Duck hepatitis B virus (DHBV).
OC Viruses; Retrovird viruses; Hepadnaviridae; Avihepadnavirus.
OX NCBI_TaxID=12639;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=INDIAN;
RA Wang C.-Y.J., Giambrone J.J., Dormitorio T.V.;
RT "The complete sequence of Duck Hepatitis B virus Indiana isolate.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF493985; AAM11781.1; -.
SQ SEQUENCE 328 AA; 36146 MW; BE6D1C9E73FA1556 CRC64;

Query Match 98.4%; Score 857; DB 12; Length 328;
Best Local Similarity 98.8%; Pred. No. 3e-67;
Matches 159; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MQGHPAKSMVRRIGEGELLNQLAGRMIPKGTLTWSGKFFPTLDHVDHVQMEINTLQ 60
DB 1 MQGHPAKSMVRRIGEGELLNQLAGRMIPKGTLTWSGKFFPTLDHVDHVQMEINTLQ 60
QY 61 NOGAMPAGARRVGLSNPTTPOIQTPEEDQKAREAFRRYQERPPETTTIPSSPP 120
DB 61 NOGAMPAGARRVGLSNPTTPOIQTPEEDQKAREAFRRYQERPPETTTIPSSPP 120
QY 121 QWKLPQDDPLNGSLLETHPLYSQSEPAVPVTKPPLKKK 161
```

```
DB 121 QWKLPQDDPLNGSLLETHPLYSQSEPAVPVTKPPLKKK 161

RESULT 4
O72885
ID O72885 PRELIMINARY; PRT; 330 AA.
AC O72885;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Surface protein.
OS Duck hepatitis B virus (DHBV).
OC Viruses; Retrovird viruses; Hepadnaviridae; Avihepadnavirus.
OX NCBI_TaxID=12639;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AUSTRALIAN DHBV;
RA Triyatni M.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AUSTRALIAN DHBV;
RX MEDLINE-21102973; PubMed-11161276;
RA Triyatni M., Ey P.L., Tran T., Le Mire M., Qiao M., Burrell C.J.,
RA Jilbert A.R.;
RT "Sequence comparison of an Australian duck hepatitis B virus strain
RT with other avian hepadnaviruses.";
RL J. Gen. Virol. 82:373-378(2001).
DR EMBL; AJ006350; CAA06988.1; -.
DR InterPro; IPR000349; Hepvir_surfa.
DR Pfam; PF00695; vmsa; 1.
SQ SEQUENCE 330 AA; 36789 MW; EDED4F42373ADA99 CRC64;

Query Match 90.2%; Score 785.5; DB 12; Length 330;
Best Local Similarity 88.9%; Pred. No. 5.4e-61;
Matches 144; Conservative 10; Mismatches 7; Indels 1; Gaps 1;

QY 1 MQGHPAKSMVRRIGEGELLNQLAGRMIPKGTLTWSGKFFPTLDHVDHVQMEINTLQ 60
DB 1 MQGHPAKSMVRRIGEGELLNQLAGRMIPKGTLTWSGKFFPTLDHVDHVQMEINTLQ 60
QY 61 NOGAMPAGARRVGLSNPTTPOIQTPEEDQKAREAFRRYQERPPETTTIPSSPP 120
DB 61 NOGAMPAGARRVGLSNPTTPOIQTPEEDQKAREAFRRYQERPPETTTIPSSPP 120
QY 121 QWKLPQDDPLNGSLLETHPLYSQSEPAVPVTKPPLKKK 161
DB 121 QWKLPQDDPLNGSLLETHPLYSQSEPAVPVTKPPLKKK 162

RESULT 5
Q91HP5
ID Q91HP5 PRELIMINARY; PRT; 329 AA.
AC Q91HP5;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Pres protein.
OS Duck hepatitis B virus (DHBV).
OC Viruses; Retrovird viruses; Hepadnaviridae; Avihepadnavirus.
OX NCBI_TaxID=12639;
RN [1]
RP SEQUENCE FROM N.A.
RA Hu J., Tang N., Huang A.;
RT "Sequence Analysis of a Cloned Duck Hepatitis B Virus Genome from
RT Chongqing Brown Duck.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF404406; AAK85437.1; -.
DR InterPro; IPR000349; Hepvir_surfa.
DR Pfam; PF00695; vmsa; 1.
SQ SEQUENCE 329 AA; 36436 MW; FIDFE48192CE9F97 CRC64;
```



```

Query Match      89.5%; Score 779.5; DB 12; Length 330;
Best Local Similarity 88.3%; Pred. No. 1.8e-60;
Matches 143; Conservative 10; Mismatches 8; Indels 1; Gaps 1;

QY      1 MGQHPAKSMDDRIGEGEILLNOLAGRMIPGKTWTWSGKPTIDHLVDHVTMEINTLQ 60
||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MGQOPAKSMDDRIGEGLNLNLAGRMIPGVITWSGKPTIDHLVDHVTMEVNTLQ 60
||||| ||||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY      61 NOGAWPAGAGRRRVLGNPTPOEIQPOWTPPEEQKAREAFRRYQEERPPETTTIPPSSPP 120
||||| ||||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      61 QQCAWPAGAGRRRLGNPTPHETPQOWTPEEDQKAREAFRRYQEERPPETTTIAPTST 120
||||| ||||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY      121 ONKLOPGDDPLLNQSILLETPLHYQS-EPAVPVIKTPPLKK 161
||||| ||||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      121 PWKLOPGDDPLLENKSILLETHPLYQNPEPAVPVIKTPPLKK 162
||||| ||||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 7
Q66404
ID Q66404 PRELIMINARY; PRT; 366 AA.
AC O66404;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Pre-S/S.
OS Duck hepatitis B virus (DHBV).
OC Viruses; Retroid viruses; Hepadnaviridae; Avihepadnavirus.
OX NCBI_TaxID=12639;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DHBVQA34;
RA Tong S., Mattes F., Blum H.E., Fernholz D., Schneider R., Will H.;
RT *complete nucleotide sequence of a chinese Hepatitis B virus.*;
RL Submitted (JUN-1991) to the EMBL/GenBank/DDBJ databases.
```

Tong S., Mattes F., Blum H.E., Fernholz D., Schneider R., Will H.; RA  
 "Complete nucleotide sequence of a chinese Hepatitis B virus."; RT  
 Submitted (JUN-1991) to the EMBL/GenBank/DBJ databases.

```

AC Q9WFA9;
AD 01-NOV-1999 (TrEMBLrel. 12, Created)
AE DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
AF DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
AG DE Pres antigen.
AH OS snow goose hepatitis B virus.
AI OC Viruses; Retrovird viruses; Hepadnaviridae; Avihepadnavirus.
AJ NCBI_TaxID=89623;
AK [1]
AL SEQUENCE FROM N.A.
AM RP STRAIN=SGHBV1-19;
AN RC MEDLINE=99420377; PubMed=10489339;
AO RA Chang S.F., Netter H.J., Bruns M., Schneider R., Frolich K., Will H.;
AP RT "A new avian hepadnavirus infecting snow geese (Anser caerulescens) produces a significant fraction of virions containing single-stranded DNA.";
AQ RL Virology 262:39-54(1999).
AR RN [2]
AS SEQUENCE FROM N.A.
AT RP STRAIN=SGHBV1-19;
AU RC Chang S.-F., Netter H.J., Bruns M., Schneider R., Froelich K., Will H.;
AV RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
AW DR EMBL; AFIL0998; AAD21992.1; -.
AX DR InterPro; IPR000349; Hepvir_surfaG.
AY DR Pfam; PF00695; VMGA; 1.
AZ SQ SEQUENCE 329 AA; 36645 MW; D896E4B2979F201E CRC64;

Query Match      83.1%; Score 724; DB 12; Length 329;
Best Local Similarity 85.2%; Pred. No. 1.3e-55;
Matches 138; Conservative 8; Mismatches 14; Indels 2; Gaps 2;

QY 1 MGQHAPKSMQVRRTTEGEIILLNLQAGRMIPKGTITWSGKPTLDHVLDDHVOTMEEINTLQ 60
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1 MGHQAKSMQARRIEGGIILLNLQAGRMIPGTVTWSGKFPSIDHVMHDVOTMEEINTLQ 60
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 61 NOGAWPAGAGRRRVGLSNPTPEIQPPQQTWEEDOKAREAFRRYQEERPPETTTTPSSPP 120
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 61 KOGAPEGAGRRRVGLTNPTPEIQPPHWTPEEDQKAREAFRRYQEERPETTTIPP-TPT 119
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 121 QWKLPQGDDPLNGSLLETHPLYQ-SEPAVPVIKTPLKKK 161
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 120 PWKLPGQDDPLLETGKSULETRLOQNSEPAPVPIKPLVKKK 161
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 10
Q9WFB3 PRELIMINARY; PRT; 329 AA.
ID Q9WFB3
AC Q9WFB3;
AD 01-NOV-1999 (TrEMBLrel. 12, Created)
AE DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
AF DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
AG DE Pres antigen.
AH OS snow goose hepatitis B virus.
AI OC Viruses; Retrovird viruses; Hepadnaviridae; Avihepadnavirus.
AJ NCBI_TaxID=89623;
AK [1]
AL SEQUENCE FROM N.A.
AM RP STRAIN=SGHBV1-7;
AN RC MEDLINE=99420377; PubMed=10489339;
AO RA Chang S.F., Netter H.J., Bruns M., Schneider R., Frolich K., Will H.;
AP RT "A new avian hepadnavirus infecting snow geese (Anser caerulescens) produces a significant fraction of virions containing single-stranded DNA.";
AQ RL Virology 262:39-54(1999).
AR RN [2]
AS SEQUENCE FROM N.A.
AT RP STRAIN=SGHBV1-7;
AU RC Chang S.-F., Netter H.J., Bruns M., Schneider R., Froelich K., Will H.;
AV RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
AW DR EMBL; AFIL0999; AAD21997.1; -.
AX DR InterPro; IPR000349; Hepvir_surfaG.

```

```
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Pres antigen.
OS snow goose hepatitis B virus.
OC Viruses; Retroid viruses; Hepadnaviridae; Avihepadnavirus.
OX NCBI_TaxID=89623;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=SGHBV1-15;
RX MEDLINE=99420377; PubMed=10489339;
RA Chang S.-F., Netter H.J., Bruns M., Schneider R., Froelich K., Will H.;
RT "A new avian hepadnavirus infecting snow geese (Anser caerulescens)
RL produces a significant fraction of virions containing single-stranded
RN DNA.";
RN Virology 262:39-54(1999).
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=SGHBV1-15;
RA Chang S.-F., Netter H.J., Bruns M., Schneider R., Froelich K.,
Will H.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF110997; AAD21987.1; -
DR InterPro; IPR000349; Hepvir_surfaG.
DR Pfam; PF00695; VMSA; 1.
SQ SEQUENCE 329 AA; 36586 MW; 7C1928C4C0E97466 CRC64;

Query Match 82.0%; Score 714; DB 12; Length 329;
Best Local Similarity 84.0%; Pred. No. 9.8e-55;
Matches 136; Conservative 9; Mismatches 15; Indels 2; Gaps 2;

Qy 1 MGQHPAKSMOVRRIEGGELLNQLAGRMIPKGTWTWSGKFTPLDHLVHVDVOTMEINTLQ 60
Db 1 MGQHPAKSMOVRRIEGGELLNQLAGRMIPKGTWTWSGKFTPLDHLVHVDVOTMEINTLQ 60
Qy 61 NQAWPAGAGRRVGLSNPTQEIPOQWTPEDOKARAFRRYQERPEPTTIPPSP 120
Db 61 KQAWPEGAGRRVGLSNPTQEIPOQWTPEDOKARAFRRYQERPEPTTIPPSP 119

Query Match 82.0%; Score 714; DB 12; Length 329;
Best Local Similarity 84.0%; Pred. No. 9.8e-55;
Matches 136; Conservative 9; Mismatches 15; Indels 2; Gaps 2;

Qy 121 QWKLQPGDPLLGNSLLETHPLQ--SEPAVPVKTPLKK 161
Db 121 QWKLQPGDPLLGNSLLETHPLQ--SEPAVPVKTPLKK 161

Query Match 51.0%; Score 444; DB 12; Length 337;
Best Local Similarity 50.0%; Pred. No. 4.4e-31;
Matches 85; Conservative 29; Mismatches 40; Indels 16; Gaps 4;

Qy 1 MGQHPAKSMOVRRIEGGELLNQLAGRMIPK---GTLTWSGKFTPLDHLVHVDVOTMEIN 57
Db 1 MGHTQAKSTTDRRVEGGELLQLAGRMIPREFQPIITAGKLPISIDHVDHDSVEELR 60
Qy 58 TLQNGAWPAGAGRRVGLSNPTQEIPOQWTPEDOKARAFRRYQERPEPTTIPP- 116
Db 61 TWQNGHWPGETGRKGLDCKPTTTPPPAITWTREDEKAKOFFKOYQENRKPQNTAPPP 120
Qy 117 -----SSPPQWKLQPGDPLLGNSLLETHPLQSEPAVPVKTPLKK 160
Db 121 LPELHAADPPQWKIKPG-DPLLQAGSLIP-----KKDPDVPILKLPQLPK 164

RESULT 15
Q8UYX4
ID 08UYX4 PRELIMINARY; PRT; 337 AA.
AC 08UYX4;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Pres/surface antigen.
OS stork hepatitis B virus.
OC Viruses; Retroid viruses; Hepadnaviridae; Avihepadnavirus.
OX NCBI_TaxID=110944;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=STHBV-21;
RA Pult I.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
```

```
Qy 61 NQAWPAGAGRRVGLSNPTQEIPOQWTPEDOKARAFRRYQERPEPTTIPPSP 120
Db 61 QQAWPEGAGRRVGLSNPTQEIPOQWTPEDOKARAFRRYQERPEPTTIPPSP 120
Qy 121 --QWKLQPGDPLLGNSLLETHPLQ--SEPAVPVKTPLKK 161
Db 121 KQWELAPGD-----PLLTQPLYPAPPAEPDIPVTKPKVPKK 160

RESULT 14
Q8UYX6
ID 08UYX6 PRELIMINARY; PRT; 337 AA.
AC 08UYX6;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Pres/surface antigen.
OS stork hepatitis B virus.
OC Viruses; Retroid viruses; Hepadnaviridae; Avihepadnavirus.
OX NCBI_TaxID=110944;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=STHBV-16;
RA Pult I.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=STHBV-16;
RA Pult I., Netter H.J., Bruns M., Prassolov A., Sirna H., Hohenberg H.,
RT "Identification and analysis of a new hepadnavirus in white storks.";
RL Virology 289:114-128(2001).
DR EMBL; AJ251936; CAC80819.1; -
DR InterPro; IPR000349; Hepvir_surfaG.
DR Pfam; PF00695; VMSA; 1.
SQ SEQUENCE 337 AA; 37684 MW; DEB85829A656121E CRC64;

Query Match 51.0%; Score 444; DB 12; Length 337;
Best Local Similarity 50.0%; Pred. No. 4.4e-31;
Matches 85; Conservative 29; Mismatches 40; Indels 16; Gaps 4;

Qy 1 MGQHPAKSMOVRRIEGGELLNQLAGRMIPK---GTLTWSGKFTPLDHLVHVDVOTMEIN 57
Db 1 MGHTQAKSTTDRRVEGGELLQLAGRMIPREFQPIITAGKLPISIDHVDHDSVEELR 60
Qy 58 TLQNGAWPAGAGRRVGLSNPTQEIPOQWTPEDOKARAFRRYQERPEPTTIPP- 116
Db 61 TWQNGHWPGETGRKGLDCKPTTTPPPAITWTREDEKAKOFFKOYQENRKPQNTAPPP 120
Qy 117 -----SSPPQWKLQPGDPLLGNSLLETHPLQSEPAVPVKTPLKK 160
Db 121 LPELHAADPPQWKIKPG-DPLLQAGSLIP-----KKDPDVPILKLPQLPK 164

RESULT 15
Q8UYX4
ID 08UYX4 PRELIMINARY; PRT; 337 AA.
AC 08UYX4;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Pres/surface antigen.
OS stork hepatitis B virus.
OC Viruses; Retroid viruses; Hepadnaviridae; Avihepadnavirus.
OX NCBI_TaxID=110944;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=STHBV-21;
RA Pult I.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
```

```

RC STRAIN-STHVV-21;
RX MEDLINE=2148937; PubMed=11601923;
RA Pult I., Netter H.J., Bruns M., Prassolov A., Sirna H., Hohenberg H.,
RT Chang S.-F., Froelich K., Krone O., Kaleta E.F., Will H.;
"Identification and analysis of a new hepadnavirus in white storks.";
RL Virology 289:114-128(2001).
DR EMBL: AJ251937; CAC80822.1;
DR InterPro: IPR000349; Hepvir_surfa9.
DR Pfam: PF00695; VMSA; 1.
SQ SEQUENCE 337 AA; 37698 MW; 1EFB2292C13BCE70 CRC64;

Query Match      51.0%; Score 444; DB 12; Length 337;
Best Local Similarity 50.0%; Pred. No. 4.4e-31;
Matches 85; Conservative 29; Mismatches 40; Indels 16; Gaps 4;

QY 1 MGCHPAKSMVDVRIEGEILLNQLAGRMIPK---GTLTWSGKFPITLDHVLHVQTMEEIN 57
Db 1 MGHTQAKSTTDRRVEGGELLQQLAGRMIPREFQGPITTAGKLPISIDHVMHDHDSVEELR 60

QY 58 TLQNCWAPGACRRVGLSNPTPQEIPOQWTPPEEDQKAREAFRRYQERRPETTTIPP- 116
Db 1 TLQNCWAPGACRRVGLSNPTPQEIPOQWTPPEEDQKAREAFRRYQERRPETTTIPP- 116

QY 61 TLQNCWAPGACRRVGLSNPTPQEIPOQWTPPEEDQKAREAFRRYQERRPETTTIPP- 120
Db 1 TLQNCWAPGACRRVGLSNPTPQEIPOQWTPPEEDQKAREAFRRYQERRPETTTIPP- 120

QY 117 -----SSPQWKLQPDPLGNOSLLETHPLYOSEPAVPVKTPLKK 160
Db 117 -----SSPQWKLQPDPLGNOSLLETHPLYOSEPAVPVKTPLKK 160

QY 121 LPELHAADPPQWKIRG-DPLQAQSLIP-----RKDPDVPILKLPQLPK 164
Db 121 LPELHAADPPQWKIRG-DPLQAQSLIP-----RKDPDVPILKLPQLPK 164

```

Search completed: January 2, 2003, 13:46:52  
Job time : 49.0927 secs